



DEMANDE INTERNATIONALE PUBLIÉE EN VERTU DU TRAITE DE COOPERATION EN MATIÈRE DE BREVETS (PCT)

(51) Classification internationale des brevets ⁶ : C12N 15/53, 15/86, 15/11, 9/02, 5/10, 1/21, C12Q 1/26, 1/68, C07K 16/40, A61K 38/44, 39/395, 48/00, G01N 33/50, 33/573	A1	(11) Numéro de publication internationale: WO 98/24914 (43) Date de publication internationale: 11 juin 1998 (11.06.98)
(21) Numéro de la demande internationale: PCT/FR97/02226 (22) Date de dépôt international: 5 décembre 1997 (05.12.97) (30) Données relatives à la priorité: 96/15032 6 décembre 1996 (06.12.96) FR (71) Déposants (pour tous les Etats désignés sauf US): GENSET [FR/FR]; 24, rue Royale, F-75008 Paris (FR). INSTITUT NATIONAL DE LA SANTE ET DE LA RECHERCHE MEDICALE (INSERM) [FR/FR]; 101, rue de Tolbiac, F-75013 Paris (FR). (72) Inventeurs; et (75) Inventeurs/Déposants (US seulement): BLUMENFELD, Marta [FR/FR]; 5, rue Tagore, F-75013 Paris (FR). TCHOUMAKOV, Ilia [FR/FR]; 196, rue des Chèvrefeuilles, F-77000 Vaux le Penil (FR). GARCHON, Henri-Jean [FR/FR]; 55, rue du Ranelagh, F-75016 Paris (FR). BACH, Jean-François [FR/FR]; 180, rue de Grenelle, F-75007 Paris (FR). (74) Mandataires: MARTIN, Jean-Jacques etc.; Cabinet Regimbeau, 26, avenue Kléber, F-75116 Paris (FR).		(81) Etats désignés: AU, CA, JP, US, brevet européen (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE). Publiée <i>Avec rapport de recherche internationale.</i> <i>Avant l'expiration du délai prévu pour la modification des revendications, sera republiée si de telles modifications sont reçues.</i>
(54) Title: NUCLEOTIDE SEQUENCE CODING FOR A FLAVIN-CONTAINING MONOOXYGENASE, CORRESPONDING PROTEIN AND THEIR APPLICATIONS IN THE FIELDS OF DIAGNOSIS AND THERAPY		
(54) Titre: SEQUENCE NUCLEOTIDIQUE CODANT POUR UNE FLAVINE MONO-OXYGENASE, PROTEINE CORRESPONDANTE ET LEURS APPLICATIONS DANS LES DOMAINES DU DIAGNOSTIC ET DE LA THERAPIE		
(57) Abstract		
The invention concerns in particular the human flavin-containing monooxygenase 2 (hFMO2), and another human enzyme of the FMO, hFMOx family, their nucleotide and polypeptide sequences. The invention also concerns vectors for cloning and/or expression containing said nucleotide sequences and cells transformed by these vectors and method for preparing said polypeptides. The invention further concerns methods for selecting compounds and of diagnosing predisposition to pathologies and/or deficiencies related to FMO's and pharmaceutical compositions containing said compounds for treating and/or preventing these pathologies.		
(57) Abrégé		
La présente invention concerne notamment la flavine mono-oxygénase 2 humaine (hFMO2), ainsi qu'une autre enzyme humaine de la famille FMO, hFMOx, leurs séquences nucléotidiques et polypeptidiques. La présente invention concerne également des vecteurs de clonage et/ou d'expression contenant lesdites séquences nucléotidiques et des cellules transformées par ces vecteurs ainsi que des méthodes de préparation desdits polypeptides. L'invention comprend aussi des méthodes de sélection de composés et de diagnostic de prédisposition à des pathologies et/ou des déficiences liées aux FMOs ainsi que des compositions pharmaceutiques comportant lesdits composés destinés au traitement et/ou à la prévention de ces pathologies.		

UNIQUEMENT A TITRE D'INFORMATION

Codes utilisés pour identifier les Etats parties au PCT, sur les pages de couverture des brochures publiant des demandes internationales en vertu du PCT.

AL	Albanie	ES	Espagne	LS	Lesotho	SI	Slovénie
AM	Arménie	FI	Finlande	LT	Lituanie	SK	Slovaquie
AT	Autriche	FR	France	LU	Luxembourg	SN	Sénégal
AU	Australie	GA	Gabon	LV	Lettonie	SZ	Swaziland
AZ	Azerbaïdjan	GB	Royaume-Uni	MC	Monaco	TD	Tchad
BA	Bosnie-Herzégovine	GE	Géorgie	MD	République de Moldova	TG	Togo
BB	Barbade	GH	Ghana	MG	Madagascar	TJ	Tadjikistan
BE	Belgique	GN	Guinée	MK	Ex-République yougoslave de Macédoine	TM	Turkménistan
BF	Burkina Faso	GR	Grèce	ML	Mali	TR	Turquie
BG	Bulgarie	HU	Hongrie	MN	Mongolie	TT	Trinité-et-Tobago
BJ	Bénin	IE	Irlande	MR	Mauritanie	UA	Ukraine
BR	Brésil	IL	Israël	MW	Malawi	UG	Ouganda
BY	Bélarus	IS	Islande	MX	Mexique	US	Etats-Unis d'Amérique
CA	Canada	IT	Italie	NE	Niger	UZ	Ouzbékistan
CF	République centrafricaine	JP	Japon	NL	Pays-Bas	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norvège	YU	Yougoslavie
CH	Suisse	KG	Kirghizistan	NZ	Nouvelle-Zélande	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	République populaire démocratique de Corée	PL	Pologne		
CM	Cameroun	KR	République de Corée	PT	Portugal		
CN	Chine	KZ	Kazakhstan	RO	Roumanie		
CU	Cuba	LC	Sainte-Lucie	RU	Fédération de Russie		
CZ	République tchèque	LI	Liechtenstein	SD	Soudan		
DE	Allemagne	LK	Sri Lanka	SE	Suède		
DK	Danemark	LR	Libéria	SG	Singapour		
EE	Estonie						

Séquence nucléotidique codant pour une flavine mono-oxygénase, protéine correspondante et leurs applications dans les domaines du diagnostic et de la thérapie

La présente invention concerne notamment la
5 flavine mono-oxygénase 2 humaine (hFMO2), ainsi qu'une
autre enzyme humaine de la famille FMO, hFMOx, leurs
séquences nucléotidiques et polypeptidiques. La présente
invention concerne également, des vecteurs de clonage et/ou
d'expression contenant lesdites séquences nucléotidiques et
10 des cellules transformées par ces vecteurs ainsi que des
méthodes de préparation desdits polypeptides. L'invention
comprend aussi des méthodes de sélection de composés et de
diagnostic de prédisposition à des pathologies et/ou des
déficiences liées aux FMOs ainsi que des compositions
15 pharmaceutiques comportant lesdits composés destinés au
traitement et/ou à la prévention de ces pathologies.

Les flavines mono-oxygénases (FMOs) (Lawton et
al., 1994) forment une famille d'enzymes microsomales
catalysant l'oxydation NADPH-dépendante de nombreux
20 composés organiques exogènes (xénobiotiques) possédant un
hétéroatome nucléophile comme en particulier l'atome
d'azote, de soufre, de phosphore ou de sélénium (Ziegler,
D.M., 1988 ; Ziegler, D.M., 1993), qu'il s'agisse de
médicaments, de pesticides ou autres substances
25 potentiellement toxiques. La cystéamine est le seul
substrat endogène actuellement connu des FMOs.

Les FMOs représentent une famille multigénique.
L'expression de formes différentes de FMOs est dépendante à
la fois du tissu et de l'espèce considérés.

30 Les FMOs ont été localisées dans différents
types de tissus, en particulier le foie, les poumons et les
reins.

A ce jour, cinq isoformes des FMOs ont été
caractérisées dans l'espèce de référence, le lapin. Leur
35 homologie est de 50-60 %. Quatre de ces isoformes, FMO1,
FMO3, FMO4 et FMO5 ont été identifiées chez l'homme

(séquences GeneBank M64082, M83772, Z11737 et L37080 respectivement). Parmi les espèces mammifères, l'homologie entre FMO orthologues est supérieure à 80 %. L'existence d'une FMO2, voire d'autres isoformes, chez l'homme, peut
5 être raisonnablement postulée.

Les FMOs sont associées au réticulum endoplasmique et sont impliquées dans la détoxification de composés xénobiotiques, la mono-oxygénation permettant de transformer le xénobiotique en substance plus polaire,
10 étape préliminaire avant son excrétion. Elles peuvent également être impliquées dans l'activation métabolique de différents composés toxiques et/ou carcinogènes présents dans l'environnement.

Le mécanisme de la réaction FMO a été décrit de
15 manière détaillée (Poulsen, L.L. et al., 1995). Par opposition à toutes les autres oxydases ou mono-oxygénases connues, les FMOs ont la propriété unique de former une enzyme intermédiaire stable 4 α -hydropéroxy flavine, NADP(H)- et oxygène-dépendante, en l'absence de substrat
20 oxydable. Parce que l'énergie de catalyse est déjà présente dans l'enzyme FMO avant le contact avec son substrat potentiel, l'adéquation du substrat n'a pas besoin d'être aussi précise que pour d'autres types d'enzymes. Cette caractéristique spécifique de la FMO est responsable de la
25 grande variété de substrats acceptés par les FMOs (incluant par exemple les alkyl- et aryl-amines tertiaires et secondaires, de nombreuses hydrazines, thiocarbamides, thioamides, sulfides, disulfides, thiols).

De nombreuses molécules, composés actifs de
30 médicaments, sont reconnues comme substrats des FMOs, soit pour une N-oxydation, soit pour une S-oxydation (Gasser, 1996), parmi lesquels on trouve notamment des anti-dépresseurs, des antipsychotiques, des anti-ulcéreux, des vasodilatateurs et des anti-hypertenseurs.

35 Bien que certains substrats de FMO soient oxydés en dérivés moins actifs, de nombreux composés nucléophiles

peuvent être métabolisés en intermédiaires pouvant être plus réactifs et/ou potentiellement toxiques ; plutôt que d'être excrétés, de tels produits peuvent induire des réponses toxiques par fixation covalente à des macromolécules cellulaires, ou par d'autres mécanismes. Par exemple, les mercaptopyrimidines et les thiocarbamides peuvent être activés de manière prédominante par une activité FMO (Hines et al., 1994). De manière plus précise, il a été montré que la néphrotoxicité associée au conjugué glutathion de l'acroléine est liée à son métabolisme médié par la FMO rénale ; la FMO forme un S-oxyde qui est ensuite libéré, par réaction d'élimination catalysée en milieu basique, sous forme d'acroléine cytotoxique (Park, S.B. et al., 1992). Ainsi, les FMOs peuvent jouer un rôle important aussi bien dans les premières étapes de toxicité chimique que dans la détoxification de composés xénobiotiques.

Comme décrit ci-dessus, un grand nombre de médicaments aujourd'hui en phase d'essais cliniques, ou largement prescrits, contiennent des fonctions à caractère nucléophile de type azote, soufre, phosphore ou autres. Le rôle de la FMO dans le métabolisme oxydatif des médicaments et des composés chimiques endogènes chez l'homme est cependant mal connu.

Cashman et al. (1996) ont récemment étudié les contributions des enzymes FMOs dans le métabolisme physiologique de la cimétidine et de la S-nicotine *in vivo*. La plus grande partie de leurs résultats confirme le fait que l'activité FMO3 du foie d'adulte est responsable de l'oxygénation de la cimétidine et de la S-nicotine, cette oxygénation étant stéréospécifique. Les auteurs montrent en outre que la stéréochimie des métabolites principaux de la cimétidine et de la S-nicotine chez des petits animaux d'expérimentation est distincte de celle observée chez l'homme, et suggèrent que différentes isoformes de FMOs pouvant être prédominantes selon les espèces, ceci peut avoir des conséquences importantes quant au choix des

animaux d'expérimentation pour les programmes d'élaboration et de développement de médicaments chez l'homme.

La FMO1 est connue pour être exprimée chez l'homme dans les reins, mais pas dans le foie. La FMO2 est exprimée majoritairement dans les poumons chez toutes les espèces de mammifères testées. La FMO3 a été isolée chez l'homme dans le foie où elle est prédominante chez l'adulte. La FMO3 est l'isoforme majeure impliquée dans la sulfoxydation de la méthionine et dans l'oxygénation stéréospécifique de la cimétidine et de la S-nicotine. La FMO3 présente une spécificité pour son substrat plus grande que celle des FMO1 trouvées dans le foie de la plupart des espèces animales étudiées. La FMO4 est une isoforme mineure dont la fonction et la spécificité de substrat sont peu connues. Elle est présente dans le foie humain et est aussi exprimée dans le cerveau où elle pourrait être impliquée dans l'oxydation de médicaments antidépresseurs comme l'imipramine. La FMO5 est exprimée dans le foie de l'homme de manière moins importante que la FMO3. Son apparent manque d'efficacité en tant qu'enzyme impliquée dans le métabolisme de médicaments, suggère qu'elle pourrait être impliquée dans une fonction physiologique.

Les différents profils d'expression des isoformes de FMO selon les tissus et/ou espèces constituent donc probablement un facteur significatif contribuant aux différences d'activités FMO observées entre les tissus et/ou entre espèces.-Ainsi, la variété des formes de FMOs pourrait avoir un impact significatif sur la différence des réponses de tissus et/ou espèces à l'exposition à un composé xénobiotique. En effet, les différences observées entre les tissus et/ou espèces dans la réponse aux composés xénobiotiques et dans leur toxicité sont liées pour une part importante aux variations d'activité et de spécificité impliquées dans le métabolisme de ces substrats par les FMOs. Facteurs génétiques et spécificité tissulaire dans

l'expression des FMOs sont des facteurs importants de ces variations.

Concernant les facteurs génétiques, il a été décrit par exemple que la triméthylaminurie, pathologie présente chez 1 % de sujets blancs britanniques et qui se manifeste par une forte odeur de poisson avarié dans l'air expiré, la sueur ou l'urine, est liée à une déficience d'origine génétique du fonctionnement d'une FMO hépatique.

Pour les raisons évoquées précédemment, il existe donc aujourd'hui un besoin important d'identifier de nouvelles isoformes de FMO ainsi que les polymorphismes génétiques éventuellement associés, présentant des spécificités quant à leurs substrats et/ou leur profil d'expression tissulaire, qui pourraient être impliquées dans le métabolisme de xénobiotiques, tel que le métabolisme de médicaments ou de substances exogènes présentes dans l'environnement comme par exemple les pesticides, ou encore qui pourraient être impliquées dans une fonction physiologique. Ceci est précisément l'objet de la présente invention.

Plusieurs gènes de la famille des FMOs humaines ont été localisés sur la région 1q23-25 du chromosome 1 par hybridation *in situ* du chromosome en métaphase.

Dès lors qu'une telle région candidate a été définie, il est nécessaire d'avoir accès au fragment du génome couvrant l'intervalle où se situe(nt) le(s) gène(s) recherché(s). Cette étape passe par l'établissement d'une carte physique, à savoir le recouvrement de la région par un ensemble de fragments clonés et ordonnés. Aujourd'hui, grâce aux données de la carte intégrée CEPH/Généthon du génome humain, environ 80 % du génome est recouvert par des clones de YACs, sous clonés en BACs dont la localisation sur les chromosomes se fait par l'intermédiaire de marqueurs polymorphes et génétiquement ordonnés (Chumakov et al., 1995). Cette carte physico-génétique permet de

gagner un temps considérable, notamment par l'utilisation du séquençage exhaustif des régions d'intérêt.

Ainsi selon la présente invention, il a été établi, après localisation du BAC 123H04M sur le locus génétique 1q24-25 précédemment cité, que l'insertion qu'il porte contient les parties 3' de hFMO3, 5' de hFMO1, ainsi que la séquence complète de hFMO2, et celle d'un autre nouveau gène membre de la famille FMO, le hFMOx.

En outre, grâce à l'utilisation de banques d'étiquettes 5', on peut vérifier l'expression des gènes candidats identifiés comme précédemment : l'identification d'une étiquette hybridant à l'une des séquences candidates indique, puisque celle-ci est issue d'une banque d'ADNc, la présence d'ARNm et donc d'une expression des séquences en question dans les tissus considérés.

C'est pourquoi la présente invention concerne notamment un polynucléotide isolé, dont la séquence SEQ ID N° 1 est partiellement représentée sur la Figure 2, et qui code pour un polypeptide de séquence SEQ ID N° 3 représentée sur la Figure 1.

La présente invention concerne également un polynucléotide isolé, dont la séquence SEQ ID N° 4 est partiellement représentée sur la Figure 10, et qui code pour un polypeptide de séquence SEQ ID N° 6.

Ces deux séquences nucléotidiques sont celles de deux gènes codant pour de nouvelles enzymes de la famille des mono-oxygénases à flavine (FMO) humaines, respectivement hFMO2 et hFMOx. Ceci a été établi par comparaison des séquences identifiées aux séquences déjà connues de FMO : des homologies structurales très fortes entre les deux séquences étudiées et celles des FMO, des homologies très fortes entre la première séquence et les FMO2 connues, notamment celle de macaque (FMO2 de macaque : séquence GeneBank U59453), ainsi qu'une homologie non

suffisante de la seconde séquence avec aucune des FMO déjà répertoriées chez l'homme ont permis de conclure.

La structure exonique des gènes de la famille FMO déjà connus est entièrement conservée dans la séquence
5 nucléotidique hFMO2 selon l'invention. Les séquences de chacun des 9 exons du polynucléotide selon l'invention (Figure 3) présentent des degrés d'homologie variant de 95 % à 98 % en ADN avec la séquence correspondante de l'ARN
10 messenger de la FMO2 de macaque (Figure 4). Les divergences entre les deux séquences nucléotidiques, ainsi que leur signification envers la séquence peptidique, sont présentées sur la Figure 5. La séquence polynucléotidique
SEQ ID N° 1 selon l'invention code pour un polypeptide de séquence SEQ ID N° 3 de 535 acides aminés (Figure 1) ; la
15 séquence SEQ ID N° 2 de l'ARN messenger prédite, ainsi que la séquence polypeptidique de la protéine humaine sont homologues à 97 % avec celles de la FMO2 de macaque (Figures 6 et 7), ce qui a permis l'identification du polypeptide selon l'invention comme étant la FMO2 humaine.
20 Le polypeptide de séquence SEQ ID N° 3, représenté à la Figure 1, présente également un haut degré d'homologie avec d'autres flavines-mono-oxygénases 2 de mammifères ; ses degrés d'homologie avec d'autres protéines de la famille des flavines-mono-oxygénases sont moins forts.

25 Comme mentionné précédemment, l'absence d'homologie suffisante entre les séquences correspondant à hFMOx - séquences génomique (SEQ ID N° 4), d'ARN messenger (SEQ ID N° 5), et peptidique (SEQ ID N° 6) - et les séquences des FMOs connues, a permis de conclure qu'il s'agit d'une
30 nouvelle isoforme de FMO.

La présente invention concerne donc les séquences d'ADN ou d'ARN, l'ADN pouvant être génomique, ADN complémentaire ou synthétique, des FMOs, notamment de hFMO2 et hFMOx, ainsi que les protéines correspondantes.

35 La présente invention concerne en outre des vecteurs de clonage et/ou d'expression contenant lesdites

séquences nucléotidiques, des cellules transformées par ces vecteurs ou des animaux contenant lesdites cellules, ainsi que des méthodes de préparation desdits polypeptides sous la forme de polypeptides recombinants.

5 L'invention comprend aussi des méthodes de sélection de composé capable de moduler l'activité FMO.

L'invention concerne également des méthodes de diagnostic de prédisposition à des troubles liés à FMO ainsi que des compositions pharmaceutiques destinées au
10 traitement et/ou à la prévention de ces troubles.

Un premier exemple de tels troubles pourrait être le glaucome primaire à angle ouvert (GPAO). En effet, d'une part Sunden et al. (1996), ainsi que les inventeurs (Belmouden et al., 1996), ont identifié la région
15 chromosomique GLC1A, qui porte entre autres séquences de gènes celles connues de la famille des FMOs, en 1q23-25, comme liée à la survenue du GPAO juvénile (GPAO-J). D'autre part, un rôle possible des mono-oxygénases dans l'étiologie du glaucome a précédemment été suggéré (Schwartzman et al.,
20 1987). En effet, des métabolites de réactions d'oxydation, en inhibant l'activité Na⁺, K⁺ ATPase dans la cornée, contribueraient à la régulation de la transparence de la cornée et de la sécrétion humorale oculaire ; or, une opacité de la cornée et une hypertension oculaire sont les
25 deux critères majeurs de diagnostic de glaucome.

Ainsi, un site d'hétérozygotie, présentant une ségrégation d'ordre génotypique dans une famille étudiée pour la présence en son sein de nombreux membres atteints de GPAO-J, a été identifié par les inventeurs dans l'exon 8
30 du polypeptide hFMO2 selon l'invention.

En poursuivant la recherche des polymorphismes présents dans des populations choisies de façon adéquate, et situés dans des séquences correspondant à celles portées par l'insertion du BAC 123H04M, ou plus généralement par
35 les séquences FMOs, on pourra identifier notamment les

mutations associées aux pathologies ou troubles liés à une altération de FMO.

Les différentes isoformes des FMOs semblent moins se distinguer par la spécificité tissulaire de leur expression, que par les substrats dont elles catalysent la transformation. Comme indiqué précédemment, l'expression des FMOs a été mise en évidence dans le foie, les poumons, les reins ou le cerveau.

L'effet pathogène d'un déficit fonctionnel d'une FMO pourrait résulter d'une capacité diminuée des tissus où elle s'exprime à résister au stress oxydatif.

Plus généralement, par leur rôle dans le métabolisme oxydatif et leur fonction de détoxification, les FMOs pourraient être impliquées dans toute pathologie dégénérative ou toxique, démontrée ou à prouver, notamment celles où une mort cellulaire programmée peut être mise en évidence, et les maladies dégénératives du système nerveux central.

De façon générale, les pathologies liées au fonctionnement des FMOs, sont rassemblées sous le nom de « troubles liés à FMO ».

Parmi les troubles liés à FMO, on peut citer par exemple, mais sans s'y limiter :

- oxydation de médicaments, substrats de FMO, en dérivés moins actifs, impliquant une perte d'efficacité dudit médicament ;
- non métabolisation de médicaments actifs sous forme de métabolites, perte d'efficacité dudit médicament ;
- non métabolisation de xénobiotiques toxiques et/ou carcinogènes, dont des substances exogènes présentes naturellement dans l'alimentation, telles que des alcaloïdes végétaux, ou des substances toxiques présentes dans l'environnement, telles que les pesticides ou les herbicides ;

- métabolisation de médicaments en intermédiaires pouvant être plus réactifs, impliquant un surdosage avec possibilité d'effet secondaire ;
- métabolisation de xénobiotiques, dont les médicaments ou autres substances exogènes, en intermédiaires pouvant être potentiellement toxiques ; et/ou
- altération de la fonction physiologique dans laquelle est impliquée la FMO ; en particulier, l'altération du fonctionnement de FMO, pourrait être impliquée dans la symptomatologie du glaucome.

Par « FMO », on entendra désigner l'une quelconque des FMOs humaines connues, FMO1, FMO3, FMO4 et FMO5, ou nouvellement décrites dans la présente demande de brevet, à savoir FMO2 ou FMOx.

Certains de ces troubles pourront avoir une origine multigénique mais pour tous, les modifications d'une ou plusieurs FMOs contribuent à la survenue du trouble ou à son aggravation.

Les séquences nucléotidiques

La présente invention concerne, tout d'abord, une séquence nucléotidique isolée, caractérisée en ce qu'elle est choisie parmi :

- a) les séquences codant pour les protéines FMO2 ou FMOx humaines et leurs variants protéiques,
- b) les séquences codant pour un fragment de ces protéines et ayant au moins 10 bases,
- c) les séquences génomiques FMO2 ou FMOx humaines et leurs allèles,
- d) les séquences présentant au moins 80 %, et de préférence au moins 90 %, d'homologie avec les séquences (a) et (c),
- e) les fragments des séquences (c) ou (d) ayant au moins 10 bases,
- f) les séquences qui s'hybrident avec une séquence de (a) à (e).

Il doit être compris que la présente invention ne concerne pas les séquences nucléotidiques génomiques

dans leur environnement chromosomique naturel, c'est-à-dire à l'état naturel, il s'agit de séquences qui ont été isolées, c'est-à-dire qu'elles ont été prélevées directement ou indirectement, par exemple par copie (ADNc),
5 leur environnement ayant été au moins partiellement modifié.

Ainsi, il peut s'agir aussi bien d'ADNc que d'ADN génomique partiellement modifié ou porté par des séquences au moins partiellement différentes des séquences
10 les portant naturellement.

Ces séquences pourront également être qualifiées de « non naturelles ».

Par « séquence nucléique », on entend un fragment d'ADN et/ou d'ARN naturel isolé, ou de synthèse,
15 désignant un enchaînement précis de nucléotides, modifiés ou non, permettant de définir un fragment, un segment ou une région d'un acide nucléique.

Par « allèles », on entend désigner les séquences mutées naturelles correspondant à des polymorphismes pouvant exister chez l'être humain et,
20 notamment, ceux qui peuvent conduire au développement de troubles liés à FMO.

Par « variant protéique », on entend désigner l'ensemble des protéines mutées pouvant exister chez l'être humain, qui correspondent notamment à des troncatures, substitutions, délétions et/ou additions de résidus d'amino-acides, ainsi que les variants artificiels qui
25 seront néanmoins également appelés « variants protéiques ». Dans le cas présent, les variants sont liés en partie à la survenue de troubles liés à FMO.
30

Selon l'invention, les fragments de séquences nucléiques peuvent notamment coder pour des domaines de la protéine ou bien être utilisés comme sonde ou comme amorce dans des procédés de détection ou d'identification ou
35 d'amplification. Ces fragments présentent une taille

minimale de 10 bases et on préférera des fragments de 20 bases, et de préférence 30 bases.

Selon l'invention, l'homologie est uniquement de type statistique, elle signifie que les séquences
5 présentent au minimum 80 %, et préférentiellement 90 %, de nucléotides en commun.

Pour ce qui concerne les séquences (f), les conditions d'hybridation doivent permettre, selon l'invention, d'assurer au moins 95 % d'homologie.

10 Plus particulièrement, la présente invention concerne une séquence nucléotidique choisie parmi :

- a) les séquences codant pour un polypeptide comprenant les amino-acides selon la séquence SEQ ID N° 3, ou selon la séquence SEQ ID N° 6,
- 15 b) les séquences nucléiques de SEQ ID N° 1 ou N° 2, ou les séquences nucléiques représentées Figures 2 et 1, ou les séquences nucléiques de SEQ ID N° 4 ou N° 5, ou les séquences nucléiques représentées Figure 10, ou les séquences nucléiques codant pour les polypeptides
20 correspondants,
- c) un fragment d'une séquence selon (a) ou (b) comportant au moins 10 bases, et
- d) une séquence qui comporte par rapport aux séquences (a), (b) ou (c) au moins une mutation ponctuelle,
- 25 e) une séquence complémentaire des séquences (a), (b), (c) ou (d).

La Figure 1 représente la séquence SEQ ID N° 3, la Figure 2 représente partiellement la séquence SEQ ID N° 1 de FMO2, la Figure 10 représente partiellement la
30 séquence SEQ ID N° 4 de FMOx, telles qu'elles ont été séquencées sur un génome d'un individu ne présentant pas de troubles FMO visibles.

La structure du gène de hFMO2 est identifiée dans la Figure 3.

35 La séquence SEQ ID N° 4 de FMOx est partiellement représentée sur la Figure 10.

Pour ce qui concerne les remarques particulières sur (a), (b), (c), (d) et (e), les remarques précédentes s'appliquent.

L'invention concerne également des fragments de ces séquences, en particulier des séquences codant pour des polypeptides ayant gardé tout ou partie de l'activité de la protéine FMO.

Certaines de ces séquences peuvent être identifiées en se reportant notamment à la Figure 3 qui schématise l'organisation de hFMO2.

Ces séquences partielles peuvent être utilisées pour de nombreuses applications, comme cela sera décrit ci-après, notamment pour effectuer des constructions protéiques de type FMO ou de types différents, mais également pour réaliser par exemple des protéines FMO-like.

Si les séquences décrites sont, en général, les séquences normales, l'invention concerne également les séquences mutées dans la mesure où elles comportent au moins une mutation ponctuelle et de préférence au plus 10 % de mutation.

De préférence, la présente invention concerne des séquences nucléotidiques mutées dans lesquelles les mutations ponctuelles sont non muettes, c'est-à-dire qu'elles conduisent à une modification de l'acide-amino codé par rapport à la séquence normale. De façon encore préférée, ces mutations portent sur des acides-amino qui structurent les protéines FMO ou les fragments correspondants de celles-ci, notamment dans les régions correspondant aux sites catalytiques, aux sites régulateurs ou aux sites de fixation des cofacteurs ; les mutations peuvent également porter sur les séquences impliquées dans le transport et l'adressage ; elles peuvent aussi en particulier supprimer les cystéines ou, au contraire, en faire apparaître, mais également changer le caractère de la protéine, soit sur le plan de la charge, soit sur le plan de l'hydrophobicité.

La présente invention concerne également les mutations pouvant intervenir dans les séquences promotrices et/ou régulatrices des gènes FMO humains, lesquelles peuvent avoir des effets sur l'expression de la protéine, notamment sur son taux d'expression.

De façon générale, la présente invention s'intéresse aussi bien aux protéines FMO normales qu'aux protéines FMO mutées, ainsi qu'à leurs fragments et aux séquences d'ADN et d'ARN correspondantes.

Parmi les fragments nucléotidiques pouvant être intéressants, notamment pour le diagnostic, il faut citer également les séquences génomiques introniques du gène FMO, par exemple les séquences jonctions entre les introns et les exons.

L'invention comprend les séquences nucléotidiques selon l'invention, caractérisées en ce qu'elles comprennent au moins la mutation G.1263mac.A, telle qu'elle sera définie ci-après dans les exemples.

L'invention comprend également les séquences nucléotidiques selon l'invention, caractérisées en ce qu'elles comportent au moins 10 bases ainsi que lesdites séquences nucléotidiques, utilisables notamment comme amorce spécifique d'un allèle.

L'invention comprend également les séquences nucléotidiques selon l'invention, utilisables notamment comme amorce nucléique, de préférence caractérisées en ce que lesdites séquences sont choisies parmi les séquences SEQ ID N° 7, SEQ ID N° 8, SEQ ID N° 9 et SEQ ID N° 10.

L'invention concerne en outre les séquences nucléotidiques selon l'invention, utilisables notamment comme sonde spécifique d'un allèle, de préférence caractérisées en ce que lesdites séquences sont choisies parmi les séquences SEQ ID N° 11, SEQ ID N° 12, SEQ ID N° 13 et SEQ ID N° 14.

L'invention a également pour objet les séquences nucléotidiques selon l'invention, caractérisées en ce que lesdites séquences codent pour l'un des domaines de FMO.

Les polypeptides codés par les séquences
5 nucléotidiques selon l'invention, notamment les polypeptides de séquence SEQ ID N° 3 ou SEQ ID N° 6, font bien entendu partie de l'invention.

Dans la présente description, les termes de protéine, polypeptide ou peptide sont interchangeables.

10 La présente invention concerne l'ensemble des amorces qui peuvent être déduites des séquences nucléotidiques précédentes et qui peuvent permettre de les mettre en évidence en utilisant une méthode d'amplification telle que la méthode PCR.

15 La présente invention concerne également les séquences nucléotidiques qui peuvent comporter des nucléotides non naturels, notamment des nucléotides soufrés ou de structure α ou β .

Enfin, la présente invention concerne, bien
20 entendu, aussi bien les séquences ADN qu'ARN, ainsi que les séquences qui s'hybrident avec elles, de même que les ADN double brin correspondants.

Parmi les fragments d'acides nucléiques intéressants, il faut citer en particulier les oligo-
25 nucléotides anti-sens, c'est-à-dire dont la structure assure, par hybridation avec la séquence cible, une inhibition de l'expression du produit correspondant. Il faut également citer les oligonucléotides sens qui, par interaction avec des protéines impliquées dans la
30 régulation de l'expression du produit correspondant, induiront soit une inhibition, soit une activation de cette expression.

Comme cela sera décrit ci-après, pour certaines applications, il peut être nécessaire de prévoir des
35 constructions mixtes, protéine/ADN/composé chimique, notamment l'utilisation d'agents intercalants par exemple ;

il doit être compris que de tels composés sont couverts par le brevet comme comportant une séquence selon l'invention.

Les protéines et polypeptides

La présente invention concerne également les
5 protéines, polypeptides ou peptides correspondant aux
séquences mentionnées précédemment, sous forme non
naturelle, c'est-à-dire qu'elles ne sont pas prises dans
leur environnement naturel mais qu'elles ont pu être
obtenues par purification à partir de sources naturelles ou
10 bien obtenues par recombinaison génétique, comme cela sera
décrit ci-après.

L'invention concerne également les mêmes
polypeptides ou protéines obtenus par synthèse chimique et
pouvant comporter des amino-acides non-naturels.

15 La présente invention concerne les protéines
recombinantes ainsi obtenues aussi bien sous forme
glycosylée que non glycosylée et pouvant présenter ou non
la structure tertiaire naturelle.

Les vecteurs et les cellules

20 La présente invention concerne également des
vecteurs de clonage et/ou d'expression comportant une
séquence nucléotidique telle que décrite précédemment.

Ces vecteurs de clonage et d'expression pourront
comporter des éléments assurant l'expression de la séquence
25 dans une cellule hôte, notamment des séquences promotrices
et des séquences de régulation efficaces dans ladite
cellule.

Le vecteur en cause pourra être à répllication
autonome ou bien destiné à assurer l'intégration de la
30 séquence au sein des chromosomes de la cellule hôte.

Dans le cas de systèmes à répllication autonome,
en fonction de la cellule hôte, procaryote ou eucaryote, on
utilisera de préférence des systèmes de type plasmidique ou
des systèmes viraux, les virus vecteurs pouvant être
35 notamment des adénovirus (Perricaudet et al., 1992), des
rétrovirus, des poxvirus ou des virus herpétiques (Epstein

et al., 1992). L'homme de métier connaît les technologies utilisables pour chacun de ces virus.

Ainsi, il est connu d'utiliser comme vecteur viral des virus défectifs dont la culture est effectuée
5 dans des cellules de complémentation, ceci évitant les risques éventuels de prolifération d'un vecteur viral infectieux.

Lorsque l'on souhaitera l'intégration de la séquence dans les chromosomes de la cellule hôte, il sera
10 nécessaire de prévoir de part et d'autre de la séquence nucléotidique à intégrer une ou plusieurs séquences provenant de la cellule hôte afin d'assurer la recombinaison. Il s'agit là également de procédés qui sont largement décrits dans la technique antérieure. On pourra,
15 par exemple, utiliser des systèmes de type plasmidique ou viral ; de tels virus seront, par exemple, les rétrovirus (Temin 1986) ou les AAV, Adenovirus Associated Virus (Carter 1993).

L'invention concerne également les cellules
20 procaryotes ou eucaryotes transformées par un vecteur tel que décrit précédemment et ceci afin d'assurer l'expression d'une protéine FMO naturelle ou variante ou bien, par exemple, d'un de ses domaines.

Les animaux, caractérisés en ce qu'ils
25 contiennent une cellule transformée selon l'invention, font également partie de l'invention.

L'invention comprend en outre un procédé de production d'un polypeptide selon l'invention, caractérisé en ce qu'on cultive une cellule selon l'invention et en ce
30 que l'on récupère la protéine produite.

Comme cela a été indiqué précédemment, la présente invention concerne également les polypeptides, obtenus par culture des cellules ainsi transformées et récupération du polypeptide exprimé, ladite récupération
35 pouvant être effectuée de façon intracellulaire ou bien de façon extracellulaire dans le milieu de culture lorsque le

vecteur a été conçu pour assurer la sécrétion du polypeptide par le biais, par exemple, d'une séquence « leader », la protéine étant exprimée sous forme d'une pré-protéine ou pré-pro-protéine. Les constructions
5 permettant la sécrétion des polypeptides sont connues, aussi bien pour des systèmes procaryotes que pour des systèmes eucaryotes. Dans le cadre de la présente invention, certains des polypeptides FMO pourront comporter leur propre système de sécrétion ou d'insertion
10 membranaire.

De préférence, l'invention concerne les polypeptides spécifiques de formes mutées des protéines selon l'invention, caractérisés en ce que leur séquence est choisie parmi les séquences polypeptidiques comprenant au
15 moins une mutation.

Parmi les cellules utilisables pour la production de ces polypeptides, il faut citer bien entendu les cellules bactériennes (Olins et Lee, 1993), mais également les cellules de levure (Buckholz, 1993), de même
20 que les cellules animales, en particulier les cultures de cellules de mammifère (Edwards et Aruffo, 1993) mais également les cellules d'insectes dans lesquelles on peut utiliser des procédés mettant en oeuvre des baculovirus par exemple (Luckow, 1993).

25 Les cellules ainsi obtenues peuvent permettre de préparer des polypeptides naturels ou variants FMO, mais également des fragments de ces polypeptides, notamment des polypeptides pouvant correspondre aux différents domaines en cause.

30 L'invention comprend également les anticorps mono- ou polyclonaux dirigés contre les polypeptides selon l'invention, de préférence, caractérisés en ce qu'ils sont obtenus par réaction immunologique d'un organisme humain ou animal avec un agent immunogène constitué par un
35 polypeptide selon l'invention, notamment un polypeptide recombinant ou synthétique selon l'invention ; de

préférence l'agent immunogène sera constitué par un polypeptide spécifique de la forme mutée de la protéine obtenue selon le procédé précédemment décrit, la séquence dudit polypeptide étant choisie parmi les séquences polypeptidiques comprenant au moins une mutation.

L'invention concerne également les anticorps selon l'invention, caractérisés en ce qu'il s'agit d'anticorps marqués, notamment pour l'imagerie.

Ces anticorps monoclonaux ou polyclonaux marqués et correspondant notamment à tout ou partie des protéines mutées pourront être utilisés par exemple comme agent d'imagerie, *in vivo* ou *ex vivo* sur des prélèvements biologiques (imagerie à l'aide d'anticorps couplés à une molécule détectable en imagerie de type PET-scan, par exemple).

Les modèles cellulaires

Les cellules transformées telles que décrites précédemment pourront également être utilisées à titre de modèle afin d'étudier les interactions entre les FMOs et les partenaires, composés chimiques et protéiques, impliqués directement ou indirectement dans l'activité FMO, et afin d'étudier les différentes interactions mises en cause selon qu'il s'agit d'une FMO normale ou d'un variant. Mais surtout ils pourront être utilisés pour la sélection de produits interagissant avec les FMOs, normales ou variantes, à titre d'agoniste, notamment d'activateur enzymatique, ou d'antagoniste, notamment d'inhibiteur enzymatique.

Une autre application potentielle de la caractérisation de ces gènes est donc la possibilité d'identifier des composés, notamment protéiques, interagissant avec ces FMOs. Il peut s'agir aussi bien d'inhibiteurs que d'activateurs, de substrats ou de cofacteurs, par exemple. Leur identification permettra de les utiliser en fonction de leurs interactions avec la protéine normale ou la protéine variante. En particulier,

on pourra chercher à isoler des agents ayant des effets différents sur les FMOs normales et variantes.

On pourra aussi utiliser ces modèles cellulaires pour étudier le métabolisme de xénobiotiques, médicaments
5 ou autres, par une FMO, normale ou variante. Ceci pourra être mis en oeuvre dans l'identification du pouvoir toxique de certains composés, dans la sélection et le développement de composés à toxicité réduite, ou à activité accrue, ou dans celui de FMOs modifiées, ayant un meilleur pouvoir de
10 métaboliser les composés d'intérêt.

Ce type de modèle cellulaire peut être réalisé en mettant en oeuvre des techniques de génie génétique. Il s'agit, suivant le type de cellules que l'on désire utiliser, de cloner le gène en question sous sa forme
15 normale ou sous sa forme mutée dans un vecteur d'expression, qu'il s'agisse d'un vecteur à réplication autonome ou d'un vecteur d'intégration, ledit vecteur comportant l'ensemble des éléments permettant l'expression du gène dans la cellule en cause, ou celle-ci ayant
20 l'ensemble des éléments permettant l'expression de la séquence en cause.

On obtient ainsi des cellules eucaryotes ou procaryotes exprimant la ou les protéines FMO, normales ou variantes, qui pourront alors constituer des modèles
25 permettant de tester tout à la fois les interactions de différents produits avec les protéines FMO ou leurs variants, ou de tester des composés, notamment des produits chimiques de synthèse, pouvant interagir avec le produit du gène FMO, normal ou muté, et ce en les ajoutant dans le
30 milieu de culture desdites cellules.

Il faut, en particulier, remarquer que les produits en question pourront aussi bien être des agents à activité antagoniste qu'agoniste.

L'utilisation de modèles cellulaires en vue de
35 tester des composés pharmaceutiques est bien connue, là encore il n'y a pas lieu de détailler ce type de modèle. On

peut cependant citer, parmi les techniques utilisées, le « Phage Display » (Allen et al., 1995), et les méthodes de double-hybride (Luban et Goff., 1995).

Ces modèles peuvent être de type *in vitro*, par exemple des cultures de cellules humaines, soit en culture normale, soit éventuellement sous forme d'organe isolé.

La présente invention concerne également des organismes tels que les animaux, en particulier des souris, exprimant le phénotype correspondant à la FMO normale ou variante d'origine humaine. Là encore, ces animaux pourront être utilisés comme animaux modèles pour tester l'efficacité de certains produits pharmaceutiques.

La présente invention concerne également les produits obtenus par la mise en oeuvre des modèles cellulaires précédents.

Méthode de diagnostic

La présente invention concerne, comme cela a été dit précédemment, plus particulièrement des méthodes de diagnostic de prédisposition à des troubles liés à FMO chez un patient, caractérisées en ce qu'on détermine à partir d'un prélèvement biologique dudit patient la présence d'une mutation dans au moins une séquence codant pour une FMO par l'analyse de tout ou partie d'une séquence nucléique correspondant audit gène, la présence d'au moins une telle mutation étant indicative d'une prédisposition dudit patient à des troubles liés à FMO.

Il est important de préciser que la présente invention ne décrit en détail que hFMO2 et hFMOx, mais les méthodes de diagnostic et les compositions à visées thérapeutiques concernent aussi bien les FMOs précédentes que FMO1, FMO3, FMO4 et FMO5. En effet, les FMOs en général interviennent dans le métabolisme des xénobiotiques et les troubles qui y sont associés, tels que, par exemple, les xénobiotiques et les troubles liés à FMO cités précédemment.

Parmi les mutations qui sont recherchées, il faut citer plus particulièrement la mutation G.1263mac.A. (localisée sur la Figure 6).

Les séquences d'acides nucléiques analysées
5 pourront être aussi bien de l'ADN génomique, un ADNc ou un ARNm.

Comme cela a été dit précédemment, parmi les troubles liés à FMO qui peuvent être mis en évidence, on entend plus particulièrement les pathologies associées au métabolisme de xénobiotique telles que citées précédemment,
10 ou associées à la fonction biologique de FMO, mais il peut exister d'autres troubles qui pourraient être liés à une anomalie des FMOs.

Les outils de diagnostic basés sur la présente invention, bien qu'ils puissent permettre un diagnostic positif et différentiel chez un patient pris isolément, seront de préférence intéressants pour un diagnostic présymptomatique chez un sujet à risque, notamment avec antécédent familial, et il est possible également de
20 prévoir un diagnostic anté-natal.

En outre, la mise en évidence d'une mutation spécifique peut permettre un diagnostic évolutif, notamment quant à l'intensité du trouble ou à l'époque probable de son apparition.

25 Les méthodes permettant de mettre en évidence la mutation dans un gène par rapport au gène naturel sont, bien entendu, très nombreuses. On peut essentiellement les diviser en deux grandes catégories, le premier type de méthode est celui dans lequel la présence d'une mutation
30 est détectée par comparaison de la séquence mutée avec la séquence correspondante naturelle non mutée, et le second type dans lequel la présence de la mutation est détectée de façon indirecte, par exemple par la mise en évidence de misappariements dus à la présence de la mutation.

35 Dans les deux cas, on préférera en général les méthodes dans lesquelles tout ou partie de la séquence

correspondant à FMO est amplifiée préalablement à la mise en évidence de la mutation, ces méthodes d'amplification pouvant être réalisées par des méthodes dites PCR ou PCR-like. Par PCR-like on entendra désigner toutes les méthodes
5 mettant en oeuvre des reproductions directes ou indirectes des séquences d'acides nucléiques, ou bien dans lesquelles les systèmes de marquage ont été amplifiés, ces techniques sont bien entendu connues, en général il s'agit de l'amplification de l'ADN par une polymérase ; lorsque
10 l'échantillon d'origine est un ARN il convient préalablement d'effectuer une transcription réverse. Il existe actuellement de très nombreux procédés permettant cette amplification, par exemple les méthodes dites NASBA « Nucleic Acid Sequence Based Amplification » (Compton
15 1991), TAS « Transcription based Amplification System » (Guatelli et al., 1990), LCR « Ligase Chain Reaction » (Landegren et al., 1988), « Endo Run Amplification » (ERA), « Cycling Probe Reaction » (CPR), et SDA « Strand Displacement Amplification » (Walker et al., 1992), bien
20 connues de l'homme du métier.

Le Tableau 1 présente des séquences d'amorces utilisables pour amplifier les séquences intéressant la mutation G.1263mac.A.

Le réactif utilisé pour détecter et/ou
25 identifier une mutation du gène FMO dans un échantillon biologique comprend une sonde dite de capture et/ou une sonde dite de détection, l'une au moins de ces sondes comportant une séquence selon la présente invention décrite précédemment.

30 Recherche de mutations ponctuelles

De façon générale, plusieurs méthodes de détection peuvent être appliquées ou adaptées si nécessaire, après amplification des séquences d'intérêt par PCR. A titre d'exemples, on peut citer :

- 1) Séquençage : comparaison des séquences de plusieurs individus et/ou repérage d'un site d'hétérozygotie chez un seul individu.
- 2) « Single nucleotide primer extension » (Syvanen et al., 1990). Des exemples d'amorces utilisables pour détecter la mutation G.1263mac.A par cette méthode figurent dans le Tableau 2.
- 3) RFLP « Restriction Fragment Length Polymorphism ». Un exemple d'enzyme de restriction utilisable pour détecter la mutation G.1263mac.A. par RFLP est présenté sur le Tableau 3.
- 4) Recherche de « Single Strand Conformation Polymorphisms » (SSCP).
- 5) Méthodes basées sur un clivage des régions misappariées (clivage enzymatique par la S1 nucléase, clivage chimique par différents composés tels que la pipéridine ou le tétrioxide d'osmium, etc.
- 6) Mise en évidence d'hétéroduplex en électrophorèse.
- 7) Méthodes basées sur l'utilisation en hybridation de sondes oligonucléotidiques spécifiques d'allèles : « Allele Specific Oligonucleotide » (ASO) (Stoneking et al., 1991). Des exemples de sondes utilisables pour la détection de la mutation G.1263mac.A. par ASO figurent sur le Tableau 4.
- 8) Méthode OLA « dual color Oligonucleotide Ligation Assay » (Samiotaki et al., 1994).
- 9) Méthode ARMS « Amplification Refractory Mutation System », ou ASA « Allele Specific Amplification », ou PASA « PCR Amplification of Specific Allele » (Wu et al., 1989).

Cette liste n'est pas exhaustive, et d'autres méthodes bien connues peuvent être utilisées.

Recherche de remaniements, par exemple de type délétions

D'autres méthodes bien connues et basées sur les techniques d'hybridation à l'aide de sondes génomiques, de sondes ADNc, de sondes oligonucléotidiques ou de ribosondes

peuvent être utilisées pour la recherche de ce type de remaniements.

Font donc, ainsi, également partie de l'invention, les méthodes de diagnostic d'une
5 prédisposition à des troubles liés à FMO chez un patient selon l'invention, caractérisées en ce que ladite analyse est réalisée par hybridation, ladite hybridation étant réalisée de préférence à l'aide d'au moins une sonde oligonucléotidique spécifique de l'allèle, ou en ce que la
10 présence d'une mutation est détectée par comparaison avec la séquence correspondante naturelle non mutée, ou en ce que ladite analyse est réalisée par séquençage, ou par migration électrophorétique, et plus particulièrement par SSCP ou DGGE, ou en ce que ladite analyse est réalisée par
15 une méthodologie visant à détecter une troncation de la protéine.

Font aussi partie de l'invention, les méthodes de diagnostic d'une prédisposition à des troubles liés à FMO chez un patient selon l'invention, caractérisées en ce
20 que tout ou partie de la séquence nucléique du gène FMO est amplifiée préalablement à la mise en évidence de la ou des mutations, de préférence l'amplification est réalisée par PCR ou PCR-like, les amorces choisies pour réaliser l'amplification étant de préférence choisies parmi les
25 amorces selon l'invention.

Les réactifs pour détecter et/ou identifier une mutation du gène FMO dans un échantillon biologique, caractérisés en ce qu'il comprennent une sonde dite de capture et/ou une sonde dite de détection, l'une au moins
30 de ces sondes comportant une séquence selon l'invention, ou un anticorps selon l'invention, font également partie de l'invention.

Méthodes basées sur la détection du produit du gène

Les mutations du gène FMO peuvent être
35 responsables de différentes modifications du produit de ce gène, modifications utilisables pour une approche

diagnostique. En effet, les modifications d'antigénicité peuvent permettre la mise au point d'anticorps spécifiques. Toutes ces modifications peuvent être utilisées en approche diagnostique, grâce à plusieurs méthodes bien connues
5 basées sur l'utilisation d'anticorps mono- ou polyclonaux reconnaissant la protéine normale ou des variants mutés, par exemple méthode RIA ou ELISA.

Enfin, il est également possible de diagnostiquer une prédisposition à des troubles liés à FMO, chez un patient, en mesurant l'activité enzymatique de la
10 (ou des) FMO à partir d'échantillons biologiques dudit patient. La mesure de cette (de ces) activité(s), par comparaison avec un étalon, interne ou externe, sera en effet indicative d'une prédisposition à l'un des troubles
15 précédemment cités.

Compositions thérapeutiques

La présente invention concerne également les traitements thérapeutiques, curatifs ou préventifs, de troubles liés à FMO.

20 On pourra utiliser les composés impliqués directement ou indirectement dans l'activité FMO, issus de l'utilisation des modèles cellulaires décrits précédemment.

On pourra particulièrement utiliser les composés capables d'interagir avec les FMOs, normales ou variantes,
25 à titre d'agoniste ou d'antagoniste notamment.

La présente invention concerne également des compositions thérapeutiques comportant à titre de principe actif un composé capable de moduler l'activité FMO, il peut s'agir de composés à activité pro-FMO, notamment tels que
30 décrits précédemment, ou des composés à activité anti-FMO.

De façon générale, on entendra par composé à « activité pro-FMO » un composé qui induira l'activité FMO, au contraire un composé anti-FMO aura tendance à réduire l'activité FMO. L'effet réel de ces types d'activités
35 dépendra du type d'enzyme exprimée, normale ou pathologique.

De façon préférée, on pourra utiliser des compositions thérapeutiques dont l'activité sera différente envers les enzymes FMOs normales et variantes.

Il est tout d'abord possible de prévoir un traitement de substitution, c'est-à-dire des compositions thérapeutiques caractérisées en ce qu'elles comportent à titre de principe actif un composé à activité pro-FMO ; il pourra s'agir notamment de tout ou partie de polypeptides tels qu'ils ont été décrits précédemment, ou bien d'un vecteur d'expression de ces mêmes polypeptides, ou bien encore de composés chimiques ou biologiques ayant une activité pro-FMO, une activité FMO-like ou induisant la production de FMO.

Il est possible également d'utiliser des compositions thérapeutiques dans lesquelles le principe actif aura une action anti-FMO, en particulier anti-FMO variante. Dans ce cas il s'agit d'un traitement suppressif. Il pourra s'agir, par exemple, de composés interagissant avec lesdites enzymes, notamment des composés protéiques, et en particulier d'anticorps anti-FMO, notamment lorsque ces anticorps reconnaîtront les protéines variantes. Il pourra s'agir également de produits chimiques ayant une activité anti-FMO, notamment des antagonistes de FMO variante.

Parmi les nombreux composés pharmaceutiques utilisables, il faut citer plus particulièrement, les séquences anti-sens interagissant avec le gène FMO normal ou muté, ou bien les séquences sens agissant sur la régulation de l'expression de ces gènes, lesdits produits pouvant également interagir en aval des produits d'expression induits par les FMOs.

Il faut également mentionner les anticorps monoclonaux inhibant les FMOs, en particulier les FMOs mutées, et/ou inhibant les ligands correspondants et/ou les produits induits par l'activité FMO, qui peuvent donc avoir des activités pro ou anti.

Il est également possible de prévoir l'expression de protéines ou leurs fragments *in vivo*, notamment par le biais de la thérapie génique et en utilisant les vecteurs qui ont été décrits précédemment.

5 Dans le cadre de la thérapie génique, il est possible également de prévoir l'utilisation des séquences des gènes ou des ADNc précédemment décrits, « nus », cette technique a notamment été développée par la société Vical, qui a montré qu'il était, dans ces conditions, possible
10 d'exprimer la protéine dans certains tissus sans avoir recours au support d'un vecteur viral notamment.

Toujours dans le cadre de la thérapie génique, il est également possible de prévoir l'utilisation de cellules transformées *ex-vivo*, lesquelles pourront être
15 ensuite réimplantées, soit telles quelles, soit au sein de systèmes de type organoïde, tel que cela est également connu dans l'état de la technique (Danos et al., 1993). On peut également envisager l'utilisation d'agents facilitant le ciblage d'un type cellulaire déterminé, la pénétration
20 dans les cellules ou le transport vers le noyau.

Ainsi, l'invention a également pour objet une composition thérapeutique, caractérisée en ce qu'elle comporte à titre de principe actif au moins un composé capable de moduler l'activité FMO, de préférence l'activité
25 FMO2 et/ou FMOx.

L'invention comprend également une composition thérapeutique caractérisée en ce qu'elle comporte à titre de principe actif au moins un composé capable d'interagir avec FMO, de préférence capable d'interagir avec FMO2 et/ou
30 FMOx, ou une composition thérapeutique selon l'invention, caractérisée en ce qu'elle présente une activité différente sur FMO normale et FMO pathologique.

L'invention comprend également une composition thérapeutique selon l'invention, caractérisée en ce qu'elle
35 comporte à titre de principe actif un composé à activité pro-FMO, de préférence choisi parmi les composés suivants :

- a) une protéine ou un polypeptide selon l'invention,
- b) un vecteur d'expression selon l'invention,
- c) une séquence nucléotidique selon l'invention, caractérisée en ce que ladite séquence est une séquence
- 5 sens induisant l'expression de FMO.

L'invention concerne en outre une composition thérapeutique selon l'invention, caractérisée en ce qu'elle comporte à titre de principe actif un composé à activité anti-FMO selon l'invention, de préférence le principe actif

10 est choisi parmi les composés suivants :

- a) un anticorps anti-FMO selon l'invention,
- b) un vecteur d'expression selon l'invention,
- c) une séquence nucléotidique selon l'invention, caractérisée en ce que ladite séquence est une séquence
- 15 antisens inhibant l'expression de FMO,
- d) une séquence nucléotidique selon l'invention, caractérisée en ce que ladite séquence est une séquence sens inhibant l'expression de FMO.

L'invention concerne aussi une composition thérapeutique selon l'invention, caractérisée en ce que le principe actif est une séquence soluble interagissant avec FMO.

20

L'invention a également pour objet l'utilisation d'un principe actif, de préférence au moins un produit selon l'invention, capable de moduler ou d'interagir avec FMO, FMO2 et/ou FMOx, pour réaliser un médicament destiné au traitement et/ou à la prévention de troubles liés au fonctionnement de FMO.

25

Sous un autre aspect, l'invention est relative à un procédé de biodégradation ou de biosynthèse de composé organique ou inorganique, caractérisé en ce qu'il met en

30 œuvre un polypeptide ou une cellule selon l'invention.

Les polypeptides à activité FMO selon l'invention pourront en effet avantageusement être utilisés pour biodégrader suivant les réactions d'oxydation, telles

35 que décrites par exemple par Ziegler (Ziegler et al.,

1993), les composés substrats de FMO, en particulier les composés tels que cités dans la présente description, ou être utilisés pour la biosynthèse de composé d'intérêt à partir desdits composés substrats de FMO, notamment pour la biosynthèse de médicament, d'additif alimentaire, de pesticide ou d'herbicide.

Les procédés d'élaboration de composé d'intérêt, caractérisés en ce qu'ils utilisent un polypeptide ou une cellule selon l'invention, font bien entendu partie de l'invention. Les polypeptides ou cellules selon l'invention, pourront en effet avantageusement être utilisés *in vitro* pour déterminer la métabolisation potentielle du composé d'intérêt et pour analyser les métabolites éventuellement obtenus, leur toxicité et/ou leur activité. Les résultats obtenus permettront de confirmer le composé ou de le reformuler de manière à ce qu'il devienne ou pas substrat de FMO, ou à ce que les métabolites formés soient différents.

Les produits susceptibles d'être obtenus par ledit procédé de biosynthèse, font également partie de l'invention.

Enfin, l'invention comprend l'utilisation de polypeptide ou de cellule selon l'invention, pour la détoxification de composé xénobiotique, substrat de FMO. Ces composés xénobiotiques peuvent être présents dans l'environnement, tels que pesticide ou herbicide, présents naturellement dans les plantes comme certains alcaloïdes, ou peuvent correspondre à des composés pharmaceutiques.

D'autres caractéristiques et avantages de la présente invention apparaîtront à la lecture des exemples ci-après, faite en se référant aux dessins annexés suivants :

- Figure 1 : Séquence polypeptidique correspondant à la séquence SEQ ID N° 3 prédite de hFMO2, homologue humaine de la FMO2 de macaque.

- Figure 2 : Séquence nucléotidique correspondant partiellement à la séquence SEQ ID N° 1 du gène codant pour hFMO2, homologue humaine de la FMO2 de macaque.

Au vu des homologies des ARN messagers connus de gènes de la famille des mono-oxygénases à flavine, ces gènes partagent la même structure exon/intron :

exon1 : non traduit, variable en taille et en séquence,
exon2 : début de la région codante, code pour les acides aminés 1-44,
exon3 : acides aminés 45-107,
exon4 : acides aminés 108-161,
exon5 : acides aminés 162-209,
exon6 : acides aminés 210-275,
exon7 : acides aminés 276-394,
exon8 : acides aminés 395-419,
exon9 : acides aminés 420-535, fin du codant et région non traduite 3'.

Les introns sont variables en taille et en complexité. Nous avons d'abord isolé la séquence de trois fragments du BAC 123H04M qui contiennent la totalité des exons de cet homologue.

Fragment 1 : contenant les exons 1 et 2,
Fragment 2 : contenant l'exon 3,
Fragment 3 : contenant les exons 4 à 9.

Les séquences de deux introns ont ensuite été complétées.

- Figure 3 :

Figure 3A : Description de la structure exon/intron du gène codant pour hFMO2, homologue humaine de la FMO2 de macaque.

Sont indiquées les positions des débuts et fins d'exons sur les séquences nucléotidiques SEQ ID N° 1 et N° 2.

Figure 3B : Description de la structure exon/intron du gène codant pour hFMOx.

Sont indiquées les positions des débuts et fins d'exons sur les séquences nucléotidiques SEQ ID N° 4 et N° 5.

- Figure 4 : Homologie entre le gène FMO2 de macaque et son homologue humain tel que présenté en figure 2.

La région 5' non traduite diverge légèrement de la séquence de macaque tel que présenté en figure 2.

- Figure 5 : Relevé des positions variantes de la séquence de l'ARNm de hFMO2 humain par rapport à la séquence homologue de macaque ; influence des variations sur la séquence protéique.

- Figure 6 : Homologies entre les séquences d'ARNm de la FMO2 de macaque et de son homologue humaine.

La position de la mutation G.1263mac.A est repérée par une flèche verticale.

- Figure 7 : Homologies entre les séquences peptidiques de la FMO2 de macaque et de son homologue humaine.

- Figure 8 : Analyse de la ségrégation du polymorphisme G.1263mac.A dans la famille étudiée.

L'ADN génomique des individus 3, 4, et 7 à 14 a été amplifié par PCR, et la séquence des fragments obtenus analysée pour détecter des sites d'hétérozygotie ségrégeant avec la maladie.

Les symboles pleins indiquent les individus atteints du GPAO juvénile. Les symboles barrés indiquent des individus non génotypés. Les individus 11 et 12 sont des jumeaux.

G/G = homozygotes pour la base en position homologue de la position 1263 de l'ARNm de la FMO2 de macaque.

G/A = hétérozygotes pour la base en position homologue de la position 1263 de l'ARNm de la FMO2 de macaque.

- Figure 9: Localisation chromosomique du BAC123H04M par hybridation in situ fluorescente.

(A) Un signal spécifique est observé sur les deux chromosomes 1. Sur la photo (B) est représenté un seul

des deux chromosomes 1. En (C), seules les bandes R de ce chromosomes sont observées, montrant que le signal de la sonde 123H04M est localisé sur la bande 1q23.

- Figure 10 : Séquence nucléotidique correspondant partiellement à la séquence SEQ ID N° 4 du gène codant pour l'isoforme hFMOx humaine.

Ce gène présente un taux d'homologie de 75 % en ADN et de 70 % en acides aminés avec les ARNs messagers de mono-oxygénases à flavine présents sur le BAC 123H04M. Il est présenté, dans cette figure préliminaire, en quatre fragments :

fragment 1 : code pour l'exon 2 (premier exon codant),
fragment 2 : exon 3,
fragment 3 : exons 4 à 8,
fragment 4 : exon 9.

EXEMPLES

Isolement du BAC 123H04M

Afin d'identifier un gène codant pour une nouvelle FMO, on a isolé un BAC (« Bacterial Artificial Chromosome ») correspondant à la région candidate précédemment localisée sur le chromosome 1. Une banque de BACs couvrant le génome humain complet a été préparée à partir de l'ADN d'une lignée lymphoblastique humaine dérivée de l'individu n° 8445 des familles du CEPH. Cette lignée a été utilisée comme source d'ADN de haut poids moléculaire. L'ADN a été partiellement digéré par l'enzyme de restriction BamH1, puis cloné au site BamH1 du plasmide pBeloBacII. Les clones ainsi obtenus ont été « poolés » et criblés selon une procédure d'analyse tridimensionnelle précédemment décrite pour le criblage des banques de YACs (« Yeast Artificial Chromosome ») (Chumakov et al., 1992). Les pools tridimensionnels obtenus ont été criblés par PCR à l'aide des amorces encadrant le marqueur D1S3423(WI-10286). Ce STS (« Sequence Tagged Site ») avait été précédemment localisé dans la région candidate. Un clone du BAC 123H04M a été ainsi isolé.

Après digestion par l'enzyme de restriction NotI, la taille de l'insert porté par ce BAC a été déterminée sur un gel d'agarose 0,8 % après migration par électrophorèse en champ alterné (CHEF) (4 heures à 9 Volts/cm, avec un angle
5 de 100°, à 11°C en tampon 0,5xTAE). On a ainsi mis en évidence que le BAC 123H04M porte un insert de 180 kb.

Localisation chromosomique du BAC 123H04M par hybridation in situ fluorescente (FISH)

La localisation chromosomique du BAC dans la
10 région candidate 1q23-q25 a été confirmée par hybridation in situ fluorescente (FISH) sur chromosomes métaphasiques, selon la méthode décrite par Cherif et al., 1990. Le BAC 123H04M a été localisé plus précisément dans la bande 1q23 du chromosome 1 (Figure 9).

15 Séquençage de l'insert du BAC 123H04M

Afin de séquencer l'insert du BAC 123H04M, on a préparé trois banques distinctes de sous clones à partir de l'ADN soniqué de ce BAC.

Après incubation une nuit, les cellules issues de
20 trois litres de culture ont été traitées par lyse alcaline selon les techniques classiques. Après centrifugation du produit obtenu dans un gradient de chlorure de césium, 52 µg d'ADN du BAC 123H04M ont été purifiés. 7 µg d'ADN ont été soniqués dans trois conditions différentes, afin d'obtenir
25 des fragments dont les tailles se distribuent uniformément de 1 à 9kb. Les fragments obtenus ont été traités dans un volume de 50 µl avec 2 unités de Vent polymérase pendant 20 minutes à 70°C, en présence des 4 déoxytriphosphates (100 µM). Les fragments aux extrémités franches résultant de
30 cette étape ont été séparés par électrophorèse en gel 1 % d'agarose à bas point de fusion (60 Volts pendant 3 heures). Les fragments groupés selon leurs tailles ont été excisés et les bandes obtenues traitées par l'agarase. Après extraction au chloroforme et dialyse sur colonnes Microcon 100, l'ADN
35 en solution a été ajusté à une concentration de 100 ng/µl. Une ligation a été effectuée, incubation une nuit, en

mettant en présence 100 ng de l'ADN fragmenté du BAC 123H04M et 20 ng d'ADN du vecteur linéarisé par digestion enzymatique, et traité par la phosphatase alcaline. Cette réaction é été réalisée dans un volume final de 10 μ l en
5 présence de 40 unités/ μ l de T4 ADN ligase (Epicentre). Les produits de ligation ont ensuite servi à transformer par électroporation, soit une souche XL-Blue (pour les plasmides multicopies), soit une souche D10HB (pour les sous clones issus du BAC). Les clones lacZ⁻ et résistants à
10 l'antibiotique, ont été repiqués individuellement en microplaques pour stockage et séquençage.

On a ainsi obtenu :

- 864 sous clones issus de l'insertion de fragments de 2 à 3 kb au site SmaI du plasmide puc18 ;
- 15 - 1728 sous clones correspondant à l'insertion de fragments de 1,5 à 2 kb au site BamHI (rendu franc) du plasmide BluescriptSK ;
- 288 sous clones portant des fragments de 4 à 7 kb insérés au site PmlI d'un vecteur BAC modifié.

20 Les inserts de ces sous clones ont été amplifiés par PCR sur cultures bactériennes incubées une nuit en utilisant les amorces des vecteurs flanquant les insertions. La séquence des extrémités de ces inserts (en moyenne 500 bases de chaque côté) a été déterminée par séquençage
25 automatique fluorescent sur séquenceur ABI 377, équipé du logiciel ABI Prism DNA Sequencing Analysis (version 2.1.2).

Les fragments de séquence provenant des sous-BACs ont été assemblés par le logiciel Gap4 de R. Staden (Bonfield et al., 1995). Ce logiciel permet la
30 reconstruction d'une séquence complète à partir de fragments de séquences. La séquence déduite de l'alignement des différents fragments est la séquence consensus.

On a enfin utilisé des techniques de séquençage dirigé (marche systématique de l'amorce) pour parfaire les
35 séquences et relier les contigs.

Analyse des séquences

Les exons potentiels du BAC 123H04M ont été repérés par recherche d'homologie sur les banques publiques de protéines, d'acides nucléiques et d'EST (Expressed Sequence Tags).

5 Banques de données :

On a utilisé des refontes locales des principales banques publiques. La banque de protéines utilisée est constituée par la fusion non redondante des banques Genpept (traduction automatique de GenBank, NCBI ; Benson et al.,
10 1996) ; Swissprot, (George et al., 1996) ; et PIR/NBRF (Bairoch et al., 1996). Les doublons ont été éliminés par le logiciel "nrdb" (domaine public, NCBI ; Benson et al., 1996). Les répétitions internes ont ensuite été masquées par le logiciel « xnu » (domaine public, NCBI ; Benson et al.,
15 1996). La banque résultante, dénommée NRPU (Non-Redundant Protein- Unique) a servi de référence pour les recherches d'homologies protéiques. Les homologies trouvées avec cette banque ont permis de localiser des régions codant potentiellement pour un fragment de protéine au moins
20 apparenté à une protéine connue (exons codants). La banque d'EST utilisée est composée des sous-sections « gbest » (1-9) de Genbank (NCBI ; Benson et al., 1996). Elle contient tous les fragments de transcrits publics.

Les homologies trouvées avec cette banque ont
25 permis de localiser des régions potentiellement transcrites (présentes sur l'ARN messager).

La banque d'acides nucléiques (autres que les EST) utilisée contient toutes autres sous-sections de Genbank et de l'EMBL (Rodriguez-Tome et al., 1996) dont les doublons
30 ont été éliminés comme précédemment.

Logiciels :

On a utilisé l'ensemble de logiciel BLAST (domaine public, Altschul et al., 1990) de recherche d'homologies entre une séquence et des banques de données protéiques ou
35 nucléiques. Les seuils de signification dépendent de la longueur et de la complexité de la région testée ainsi que

de la taille de la banque de référence. Ils ont été ajustés et adaptés à chaque analyse.

Identification de polymorphismes génétiques associés au FMO en relation avec un polymorphisme phénotypique associé à la
5 survenue du glaucome juvénile GPAO-J, maladie de
transmission autosomale dominante (locus GLC1A)

Détection de Polymorphismes/Mutations

1) Extraction de l'ADN

L'ADN est extrait du sang veineux périphérique
10 après lyse cellulaire, digestion protéique, partition organique et finalement précipitation alcoolique.

Le sang (20 ml) est prélevé par ponction veineuse périphérique sur un tube contenant de l'EDTA.

Il est dilué avec un volume d'eau bidistillée.
15 Après 10 minutes, les cellules sont collectées par centrifugation à 1600 g pendant 10 minutes. Cette manipulation est répétée.

Les cellules blanches sont lysées en présence de 20 ml de tampon CLB (Tris 10 mM pH 7.6, 5 mM $MgCl_2$, sucrose
20 0.32 M, Triton X-100 1 % (v/v). Les noyaux sont collectés par centrifugation à 1600 g pendant 10 minutes. Cette manipulation est répétée.

Les noyaux sont lavés une fois dans le tampon RSB (Tris 10 mM pH 8, NaCl 10 mM, EDTA 10 mM). Le culot est
25 resuspendu dans 2 ml de tampon RSB auquel est ajouté du lauryl sulfate de sodium (1 %) et la protéinase K (200 mg/ml). Le mélange est incubé à 55°C pendant au moins 3 heures et régulièrement agité.

La solution d'ADN ainsi obtenue est ensuite
30 extraite avec un volume de phénol équilibré avec un tampon 50 mM Tris pH 8. Cette opération est répétée et complétée par une extraction avec un volume de chloroforme /alcool isoamylique (24 : 1 v/v).

L'ADN est précipité avec un volume d'isopropanol,
35 rincé à l'ethanol (70 %), séché et enfin resuspendu dans 1 ml de tampon TE (Tris 10 mM pH 8, EDTA 0.5 mM). La

concentration d'ADN est évaluée par mesure de l'absorbance à 260 nm en utilisant l'équivalence de 50 µg/ml d'ADN pour une unité d'absorbance. La concentration d'ADN est alors ajustée à 200 µg/ml.

5 2) Amplification de l'ADN génomique

Les amorces oligonucléotidiques utilisées pour l'amplification génomique des séquences exoniques dérivées du BAC 123H04M, telles que prédites par analyse informatique, ont été définies à l'aide du logiciel OSP
10 (Hillier et al., 1991).

Toutes ces amorces contiennent, en amont des bases spécifiquement ciblées par l'amplification, une queue oligonucléotidique commune, destinée à permettre le séquençage des fragments amplifiés (PU pour les amorces en
15 amont, et RP pour les amorces en aval ; séquences exposées sur le Tableau 5).

Les amorces oligonucléotidiques ont été synthétisées selon la méthode des phosphoramidites, sur un synthétiseur GENSET UFPS 24.1.

20 L'amplification de chaque séquence exonique prédite a été réalisée par réaction d'amplification en chaîne par polymérase (PCR), dans les conditions suivantes :

Volume final	50 µl
ADN génomique	100 ng
25 MgCl ₂	2 mM
dNTP (pour chacun)	200 µM
Amorce (pour chacune)	7.5 pmoles
AmpliTaQ Gold DNA polymerase (Perkin)	1 unité
*Tampon de PCR	1 X

30 * : (10X=0.1 M Tris HCl pH 8.3, 0.5 M KCl)

L'amplification est réalisée dans un thermocycleur Perkin Elmer 9600 ou MJ Research PTC200 avec couvercle chauffant. Après un chauffage à 94°C pendant 10 minutes, 35 cycles sont effectués. Chaque cycle comprend : 30 secondes à
35 94°C, 1 minute à 55°C et 30 secondes à 72°C. Un segment

final d'élongation de 7 minutes à 72°C termine l'amplification.

La quantité de produits d'amplification obtenue est déterminée sur micro-plaque de 96 puits, par 5 fluorométrie, utilisant l'agent intercalant Picogreen (Molecular Probes).

3) Détection des polymorphismes/mutations

- Séquence

Les produits de l'amplification génomique par 10 PCR ont été séquencés sur séquenceur automatique ABI 377, en utilisant des amorces fluorescentes marquées par les fluorochromes ABI (Joe, Fam, Rox et Tamra) et l'ADN polymérase Thermosequanase (Amersham).

Les réactions ont été réalisées en microplaques 15 de 96 puits, sur thermocycleur Perkin Elmer 9600, dans des conditions classiques de cycles de température :

- 8 cycles : dénaturation : 5 sec. à 94°C; hybridation : 10 sec. ; élongation : 30 sec. à 72°C, puis
- 13 cycles : dénaturation : 5 sec. à 94°C; élongation : 30 20 sec. à 72°C.

6 unités de Thermosequanase, et 5-25 ng de produit d'amplification ont été utilisées par réaction de séquence.

A l'issue des cycles d'amplification, les produits des réactions de séquence sont précipités dans l'éthanol, 25 resuspendus dans du tampon de charge contenant de la formamide, dénaturés, et déposés sur gels d'acrylamide 4 % ; les électrophorèses (2 heures 30 à 3000 Volts) sont conduites sur séquenceurs ABI 377 équipés des logiciels ABI de collection et d'analyse (ABI Prism DNA Sequencing 30 Analysis Software, version 2.1.2.).

- Analyse des séquences

Le GPAO-J étant une maladie autosomale dominante, les données de séquence obtenues ont été analysées afin de détecter la présence de sites d'hétérozygotie parmi les 35 patients atteints de glaucome juvénile. Les sites d'hétérozygotie ont été confirmés après comparaison des

séquences des deux brins d'ADN génomique de chaque individu concerné. Un site d'hétérozygotie est retenu comme mutation candidate responsable de la survenue de troubles liés à FMO s'il est présent dans une population de membres d'une même
5 famille, alors qu'il est absent en général chez les contrôles non apparentés à la famille.

- Résultats

Parmi tous les fragments d'amplification dérivés du BAC 123H04M étudiés, l'un d'entre eux présente un site
10 d'hétérozygotie ségrégeant avec la survenue du glaucome juvénile dans un pedigree représenté sur la Figure 8.

Ce site d'hétérozygotie (G/A) est présent chez 7 patients atteints de GPAO-J tandis qu'il est absent de 3 patients sains homozygotes (G/G), tous issus de la même
15 famille. De plus, 99 contrôles non apparentés sont de même homozygotes (G/G) pour ce site, indiquant que la fréquence de l'allèle A dans la population générale est inférieure à 0.005.

Le site est contenu dans l'exon 8 du gène codant
20 pour la protéine hFMO2 selon l'invention ; la mutation décrite transforme l'acide glutamique en position 402 de la séquence SEQ ID N° 1 de la hFMO2 en lysine (Figure 1).

Il est surprenant de remarquer que le calcul des lod scores intégrant les données précédentes pour
25 différentes hypothèses de fréquences de chaque allèle dans la population générale, indique une probabilité supérieure à 100 contre 1 que l'hétérozygotie (G/A) décrite soit liée au GPAO-J (Tableau 6). Cette probabilité est significative du fait que l'analyse a porté sur une seule famille.

30 Les amorces ayant permis l'amplification du fragment d'ADN contenant ce site d'hétérozygotie sont décrites dans le Tableau 1.

Tableau 1 : Séquences des amorces utilisées pour amplifier
35 la région exonique dérivée du BAC 123H04M et contenant un site d'hétérozygotie lié au GPAO juvénile

Locus du fragment : FM02/Exon 8

Taille du fragment amplifié : 420

Amorces: Amont PU (SEQ ID N°7) : 5'TCACATAGAGTGCTATGGGGG

Aval RP (SEQ ID N°8) : 5'CTTAGGAAGAAGATAAAAATGCAAC

5 Tableau 2 : Exemples d'amorces pour détecter la mutation G.1263mac.A par « Single Nucleotide Primer Extension »

a) SEQ ID N°9 : 5' AATGTCCATCATCATAGTTCTCT 3' (antisens)
et/ou

b) SEQ ID N°10 : 5' TAGGCTTGTGTAGCCTGCCCTCA 3' (sens)

10

Tableau 3 : Identification de la mutation G.1263mac.A par RFLP

5' CCCTCAGAGAGAA 3' « normal »

15

site DdeI (C TNAG)

5' CCCTCAaAGAGAA 3' « mutant »

20

pas de coupure

Tableau 4 : Exemple de sondes pour la détection de la mutation G.1263mac.A. par la technique ASO

Spécifique de l'allèle G

25

SEQ ID N° 11 : 5' CCTCAGAGAGAACTAT 3' et sa
complémentaire :

SEQ ID N° 12 : 3' GGAGTCTCTTTGATA 5'

30

Spécifique de l'allèle A

SEQ ID N° 13 : 5' CCTCAAGAGAACTAT 3' et sa
complémentaire

35

SEQ ID N° 14 : 3' GGAGTTTCTCTTGATA 5'

Tableau 5 : Séquence des amorces utilisées pour le séquençage des fragments d'amplification à partir d'ADN génomique

5 PU 5' TGTAACGACGGCCAGT
 RP 5' CAGGAAACAGCTATGACC

10 Tableau 6 : Lod score entre le polymorphisme G.1263mac.A et le GPAO juvénile dans la famille étudiée en fonction de la fréquence des deux allèles dans la population générale.

Fréquence de l'allèle rare (A)	Θ (taux de recombinaison)	Lod score
0.01	0	2.07
0.001	0	2.10
0.0001	0	2.10
0.00001	0	2.10

SEQUENCE LISTING

SEQ ID N° 1
SEQ ID N° 2
SEQ ID N° 3
5 SEQ ID N° 4
SEQ ID N° 5
SEQ ID N° 6
SEQ ID N° 7
SEQ ID N° 8
10 SEQ ID N° 9
SEQ ID N° 10
SEQ ID N° 11
SEQ ID N° 12
SEQ ID N° 13
15 SEQ ID N° 14

RÉFÉRENCES

- Allen J.B., Walberg M.W., Edwards M.C. & Elledge S.J.
Finding prospective partners in the library: the two hybrid
5 system and phage display find a match. TIBS 20: 511-516
(1995).
- Altschul, Stephen F., Gish W., Miller W., Myers E. W., &
Lipman D.J.
10 Basic local alignment search tool.
J. Mol. Biol. 215:403-10 (1990).
- Bairoch A. & Apweiler R. The SWISS-PROT protein sequence
data bank and its new
15 supplement TREMBL. Nucleic Acids Res. 24: 21-25 (1996).
- Belmouden A., Adam M.F., Dupont de Dinechin S., Brézin
A.P., Rigault P., Chumakov I., Bach J-F., & Garchon H-J.,
1996, Recombinational and physical mapping of the locus for
20 primary open-angle glaucoma (GLC1A) on chromosome 1q23-q25.
Genomics, sous presse.
- Benson D. A., Boguski M., Lipman D. J. & Ostell J. GenBank.
Nucleic Acids Res. 24: 1-5 (1996).
25
- Bonfield J. K., Smith K. F. & Staden R. A new DNA sequence
assembly program.
Nucleic Acids Res. 23: 4992-9 (1995).
- 30 Buckholz R.G. Yeast Systems for the Expression of
Heterologous Gene Products. Curr. Op. Biotechnology 4: 538-
542 (1993).
- Cashman J.R., Park, B.P., Berkman, C.E. & Cashman, L.E.
35 Rôle of hepatic flavin-monooxygenase 3 in drug and chemical

metabolism in adult humans. *Chemico-Biological Interactions* 96: 33-46 (1995).

5 Carter B.J. Adeno-Associated virus vectors. *Curr. Op. Biotechnology* 3: 533-539 (1993).

10 Cherif D., Julier C., Delattre O., Derré J. Lathrop G.M., & Berger R.: Simultaneous localization of cosmid and chromosome R-banding by fluorescence microscopy - Applications to regional mapping of chromosome 11. *Proc.Natl.Acad.Sci. USA.* 87: 6639-6643 (1990).

15 Chumakov I., Rigault P., Guillou S., Ougen P., Billault A., Guasconi G., Gervy P., Le Gall I., Soularue P., Grinas P. et al. Continuum of overlapping clones spanning the entire human chromosome 21q. *Nature* 359: 380-386 (1992).

20 Chumakov I.M., Rignault P., Le Gall I. et al. A YAC contig map of the human genome. *Nature* 377 supplt: 175-183 (1995).

Compton J. Nucleic Acid Sequence-Based Amplification. *Nature* 350: 91-92 (1991).

25 Danos O., Moullier P. & Heard J.M. Réimplantation de cellules génétiquement modifiées dans des néo-organes vascularisés. *Médecine/Sciences* 9:62-64 (1993).

30 Edwards C.P. et Aruffo A. Current applications of COS cell based transient expression systems. *Curr. Op. Biotechnology* 4: 558-563 (1993).

Epstein A. : Les vecteurs herpétiques pour le transfert de gènes - *Médecine/Sciences* 8: 902-911 (1992).

- George D. G., Barker W. C., Mewes H. W, Pfeiffer F. & Tsugita A. The PIR-International Protein Sequence Database. Nucleic Acids Res. 24: 17-20 (1996).
- 5 Guatelli J.C. et al. Isothermal in vitro amplification of nucleic acids by a multienzyme reaction modeled after retroviral replication. Proc. Natl. Acad. Sci. USA 87: 1874-1878 (1990).
- 10 Hillier L. & Green P. OSP: a computer program for choosing PCR and DNA sequencing primers. PCR Methods Appl. 1: 124-8 (1991).
- Hines et al., Toxicol. Appl. Pharmacol. 125, 1-6 (1994).
- 15 Landegren U., Kaiser R., Sanders J. & Hood L.A ligase-mediated gene detection technique. Science 241: 1077-1080 (1988).
- 20 Lawton M.P., Cashman J.R., Cresteil T., Dolphin C.T., Elfarra A.A., Hines R.N., Hodgson E., Kimura T., Ozols J., Phillips I.R., Philpot R.M., Poulsen L.L., Rettie A.E., Shephard E.A., Williams D.E., & Ziegler D.M.: A nomenclature for the mammalian flavin-containing
- 25 monooxygenase gene family based on amino acid sequence identities. Arch. Biochem. Biophys. 308:1, 254-257 (1994).
- Luban J. & Goff S.P. The yeast two-hybride system for studying protein - protein interactions. Current Op.
- 30 Biotechnology 1995, 6:59-64.
- Luckow V.A. Baculovirus systems for the expression of human gene products. Curr. Op. Biotechnology 4: 564-572 (1993).

- Olins P.O. et Lee S.C. Recent advances in heterologous gene expression in E. coli. Curr. Op. Biotechnology 4:520-525 (1993).
- 5 Park, S.B. et al., Chem. Res. Toxicol. 5, 193-201 (1992).
- Perricaudet M., Stratford-Perricaudet L., & Briand P. : La thérapie génique par adénovirus - La Recherche 23: 471-473 (1992).
- 10 Poulsen, L.L. et al., Chem. Biol. Interact. 96, 57-73 (1995).
- Rodriguez-Tome P., Stoehr P. J., Cameron G. N., & Flores T.
- 15 P. The European Bioinformatics Institute (EBI) databases. Nucleic Acids Res. 24: 6-12 (1996).
- Samiotaki M., Kwiatkowski M. Parik J., & Landegren U. Dual-color detection of DNA sequence variants through ligase-mediated analysis. Genomics 20: 238-242 (1994).
- 20 Schwartzman, M.L., Masferrer, J., Dunn M.W., McGiff J.C., Abracham N.G., 1987, Curr Eye Res. 6 : 623-630.
- 25 Schwartzman M.L., Balazy M., Masferrer J., Abraham, N.G., McGiff, J.C., Murphy, R.C., 1987, PNAS USA 84 : 8125-8129.
- Stoneking M., Hedgecock D., Higuchi R.G., Vigilant L., & Erlich H.A. Population variation of human DNA control
- 30 region sequences by enzymatic amplification and sequence-specific oligonucleotide probes. Am. J. Hum. Genet. 48: 370-382 (1991).
- Sunden S.L.F., Alward W.L.M., Nichols B.E., Rokhlina T.R.,
- 35 Nystuen A., Stone E.M. & Sheffield V.C. Fine mapping of the autosomal dominant juvenile open angle glaucoma (GLC1A)

region and evaluation of candidate genes. Genome research
6: 862-869 (1996).

Syvänen A.C., Aalto-Setälä K., Harju L., Kontula K. , &
5 Soderlund H. A primer-guided nucleotide incorporation assay
in the genotyping of Apo E. Genomics 8: 684-692 (1990).

Temin H.M. : Retrovirus vectors for gene transfer. In
Kucherlapati R., ed. Gene Transfer, New York, Plenum Press,
10 149-187 (1986).

Walker G.T., Fraiser M.S., Schram J.L., Little M.C., Nadeau
J.G., & Malinowski D.P. Strand displacement amplification :
an isothermal in vitro DNA amplification technique.
15 Nucleic Acids Res. 20: 1691-1696 (1992).

Wu D.Y., Ugozzoli L. Pal B.K., Wallace R.B. Allele-specific
amplification of b-globin genomic DNA for diagnosis of
sickle cell anemia. Proc. Natl. Acad. Sci. USA 86: 2757-
20 2760 (1989).

Ziegler, D.M., , Drug Metab. Rev. 19, 1-32 (1988).

Ziegler, D.M., Annu. Rev. Pharmacol. Toxicol., 33, 179-199
25 (1993).

LISTE DE SEQUENCES

(1) INFORMATIONS GENERALES:

(i) DEPOSANT:

- (A) NOM: GENSET
- (B) RUE: 24 RUE ROYALE
- (C) VILLE: PARIS
- (E) PAYS: FRANCE
- (F) CODE POSTAL: 75008

(ii) TITRE DE L' INVENTION: SEQUENCE NUCLEOTIDIQUE CODANT POUR UNE
FLAVINE MONOOXYGENASE, PROTEINE CORRESPONDANTE ET LEURS
APPLICATIONS DANS LE DOMAINE DU DIAGNOSTIC ET
THERAPEUTIQUE

(iii) NOMBRE DE SEQUENCES: 14

(iv) FORME DECHIFFRABLE PAR ORDINATEUR:

- (A) TYPE DE SUPPORT: Floppy disk
- (B) ORDINATEUR: IBM PC compatible
- (C) SYSTEME D' EXPLOITATION: PC-DOS/MS-DOS
- (D) LOGICIEL: PatentIn Release #1.0, Version #1.30 (OEB)

(vi) DONNEES DE LA DEMANDE ANTERIEURE:

- (A) NUMERO DE LA DEMANDE: FR 9615032
- (B) DATE DE DEPOT: 06-AUG-1996

(2) INFORMATIONS POUR LA SEQ ID NO: 1:

(i) CARACTERISTIQUES DE LA SEQUENCE:

- (A) LONGUEUR: 26016 paires de bases
- (B) TYPE: nucléotide
- (C) NOMBRE DE BRINS: double
- (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADN (génomique)

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 1:

CATCAGTTAT CCCTGGAGGA TAACTAAGCC ATCTGCCTCC ATCATCTTTT AAGGGTTCAG	60
TCAGTTTAAA ACTTTGCTTC TATACCTAGG TATTTTCTTT TCTCTGTATG TTGGTCAGGT	120
ACAATTATTT TTAACAGGGC TTCCATCAAT ATCATAACTA CCTAGAGAAG ACATTGCAAA	180
GATAAAATTG GAGAATTGTT AACAGGCTGT TAACAAAATG TGTACCCAAC TGCCAATGAA	240
GTGGCTTGAT TTTTTCTTT TTTTAAATT TTTCTTTTGT ATCCTTTTAT TTTATTTACT	300

TATTTTTTTAG	AGACACAGTC	TCGCTCTGTT	ACCCAGGCTG	GCGTACAATG	GCACAATCAT	360
AGCCCACTGC	AGTCTCGACC	TCCAGGGCCT	CAAGCAATCC	TCCCACCTCA	GCCTTCCAAG	420
TAGCTGAGAC	TACAAGTGCA	TGCTGCCATG	CCTGACTGAT	TTTTTGTTTT	TTGCAGAGAT	480
GAAGTCTCAC	TATGTTGCCC	AGTCTGACCT	TGAACTCCTA	GCAATACCCT	ACCCTGGCCT	540
CCCAAAGTGC	TGGGATTACA	GGCATGAGCC	ACTGTGCCTG	GCCTTTTTTC	ATTTTAACTG	600
AGAAATGTGT	TCAGCTCTTT	TGTTCCCTTAG	TCATTGATCA	TCACCTTTGT	TATATCTGTT	660
AGTCTTGTCA	TAGAGTTGCT	GCACTTATTA	CACAGAGAAG	GCCTTTTATC	ACGACCAATT	720
TATTTTAGGA	AATTTAGGG	AAAACGTTTT	TCTAGAACAC	CTTATTTGAC	ATTATAAAAC	780
AACTCTTCAC	TCTTGCACTC	CAGACCTCCC	TTCCAGTTT	TCTTTTTCTC	CATAGTGGTC	840
ATCACCCTT	GTTTTATTTT	ATTGATGGGC	TGTCTGGCTC	CCTCAACTGC	AAAGTAACT	900
CCACAAAGGC	AGAGAGTTTT	GTCTCTTTTA	TTCATTGCTG	TACCTGCATC	ACTTAGAAAG	960
TTTCTGGCAC	CTAGGAAGTG	TTCAGTAAAT	ATTTATTGAA	TAAGTTTATG	TAAAACGTCT	1020
CAGACTCCTT	AGAGAACTG	GTCTTTTGGG	GTTGGAGAAT	AAAGTTCTTT	ACCTCATCAG	1080
TTAGACTCTA	TCTAAGGTAC	ACGAGGGCTT	GCTAGTCTCC	TAAGTTAGTC	TGCTAATAAA	1140
TGTTAACCCT	AATAACTGAA	ATTATTAGCA	GAGGTAATTA	TCCAGTTCTA	TATCAAGGCA	1200
AAAAGACAGC	AGTGGATAGA	AAGATCTTAG	AAGTCCCACT	AGGTTCATCC	AAGCCACCAT	1260
ACACATAGGC	AGAAAAATCA	AAATAAGATA	TGAGCCTGGA	CAGGGTGAGC	AATCTGGGAA	1320
AAGATGAACA	CAGTATGCTA	GGACCCAGAA	ATCATCAAGT	CTATGAAAAC	TAAGCCAGAA	1380
CACAAATGTG	AATTCCATAA	GATCAGGAAC	ATAATCTGTC	TTGTTTCATCC	AGGCATGGTA	1440
ATCTGCCAGA	AATAGTGCTT	AACTGCAAGA	ACTGAATATT	TGTTAGATAA	TTAAACCATC	1500
AACTAAATGA	GATTCATGCA	ACCATGAAAA	ATGCTGCTAT	AGGTACACAA	TATTGATATA	1560
CTAGAAAGTT	AAAAAATCAA	GTTGGAAATT	AGACTATTCC	ATTTCTGTTT	GTGTGTATGT	1620
ATCTACAAAT	AGGTGGAAGG	ATATACCAA	ATGTCAACAG	CAGTTACCTC	TGGGTGGTGA	1680
GGAGTAATCT	TAACCTTGTT	ATTTATCCCT	ATATGTTTAT	TTGTGAATGA	ATATTTATTA	1740
CATCATTATA	AAAAGGATTT	TTAAACTATC	TGTATGTTTA	AGAGTATATG	TTGCTACTAT	1800
GTAAGAGTAT	ATGCTGTTAC	TGTAAAGACA	TTGCATTACT	ACTGTTGACC	TCAGAGCACG	1860
CGCCTCTTGC	CTAATTCTAG	GACTCCTAAC	TAAGTCTTTG	GAGTTTCAGC	TGGAAGAATG	1920

CTGGAGGAAT	ACGGAAGTCC	TCCCATTCT	CACAGCCACC	TCCAAGTCTT	AAAAACGCTT	1980
CCAACTGCCT	CCCAGCACAC	AACCAAGGGA	GAAAGTATT	CTGTCAAAGA	GACGGTGCCA	2040
AAAGGCAAAA	ACAAAGGTAA	GGATGATCGC	TGGGGAAAGA	AGCTGAAAAG	GAAAAGCTCA	2100
GAAGTCTAGC	TGGAAATTTG	GCTCACATCC	CTAGTATGTT	ACTGCATAGT	CTGGCTTTGT	2160
TCAATGGGTC	GCTTTTAAAT	ATTAAAGCTA	GATGTAAGCA	AGGTTTGCAA	CAAAGTCCAT	2220
AAGAACTCA	GCTTTTCTCA	AAGGCAAGAA	GAGAGCAGGA	TTTTTGAAGT	GCTCTTTATT	2280
CAATAGTGCT	GCTTATTAAA	TTACCACTGC	TACAATGTTT	AAAGCCAATT	ACCTGAGCAC	2340
ATCATAAGGA	TTCTCTTACC	GGTTGTCCCA	GTAAAGTAAT	GTTGATTGAT	CAACTCCTTG	2400
ACAGGAGCTG	ATGGCAAAGA	AGGTAGCTGT	GATTGGAGCT	GGGGTCAGTG	GCCTAATTTT	2460
TCTGAAGTGC	TGTGTGGATG	AGGGACTTGA	GCCCACTTGC	TTTGAGAGAA	CTGAAGATAT	2520
TGGAGGAGTG	TGGAGGTTCA	AAGTAAGTGA	GATTTTCTTG	GGTCTTGAAC	AGGTTGTGTT	2580
GTTATTTTCA	GGTGAATCAC	AGTTACTGAT	GGGTCATATT	GAGAAATTTA	TTAAACAAGT	2640
CTGATCAGAT	TTTATTTCTA	TTTATTGATG	TGGCCATAAT	GGAAGTGAAG	TCATAGGCTG	2700
GCATCTCTCC	CCCAGTCAAT	ACTAACCCAA	CCCAGGTAGC	TGACCCAGGC	ATGTAAAAGA	2760
TCTCTTCTTT	TGGATTGAGC	AATTGTCTTA	CAGCCCATAC	TTCTGTCATT	CTTTAATACG	2820
CTAATATTAG	AGAACATTTT	ACAAAAATAG	AAGTAACAGG	GATTCTTCTC	AAGATATCAC	2880
TTCTGTTTCA	ATTATTAAAC	CAAATGCTTC	TTTAGAGACC	ATGCTCTTAT	CATTACTATT	2940
TTTCTCTGAC	AAATGAAGCA	TGTTTGTTTA	CTGAGCTTTA	TCAATGACAT	TCTAGTATAA	3000
CTGCTGTGAA	ACTCTTTGTT	AAATATGTTT	TATTAAATTT	ATTCTATTAA	TCAAACCAAA	3060
ATATTGATAA	TGCTATTTGT	CTGTATTAGT	CCATTCTCAT	GCTGCTATGA	AGAAATACTG	3120
AGACTGGGTG	ATTTATAAAG	GAAAGAGGTT	TAATTGACTC	CCAGTTCCAC	AATGCTGGGG	3180
AGGACTCAGG	AAATTTACAA	TCATGGCAGT	GGGAAAGAGA	GGTGCTGAGC	AAAGGGGGAA	3240
AAGCCCCTTA	TAAAACCATC	AGATCTCATT	AGAACGCACT	CACATCATG	AGAACAGCAT	3300
GAGGGTAGCT	GCCCCCTTGA	TTCAATTACC	TACCCCCACC	AGGTCCTTCC	CAAGACATGT	3360
GGGGATTGTG	GGAAGTACAA	TTCAATATGA	GATTTGGATG	GGGACACAAA	GCTAAACCAT	3420
GTCAGTGTCC	TTAAAAATTT	GTATAAAACT	TAGAAAGTTG	CATAGATAGC	TATAAGGAGT	3480
TACAATTATT	CCTTCCCACA	ACCTCTCAAT	AGGTAGTAGC	TTACCACCTT	CTAGCTGTGA	3540

GATCTTGAGC	AAGTTATTTA	CATCCTGTGT	TTCAATTTAC	TCAGTTATAA	ATGGATATAA	3600
TAACAGGAAA	GTGTGATTAT	CTCATAGTGC	TATTTTGAAG	ATTAAGGGAG	ATAATTCATA	3660
TAAAGAACTT	AGATAAGTTC	CGGACTCATA	GAGTTCAATA	AATGTTAGCT	ACTAATAATA	3720
ACTATATATT	TTATAGATGA	GCAAAC TGAA	AGTGAGGGAG	GTTAAGTGAG	ATGGCCAGGG	3780
CCACACAAC T	GGAGGAACTG	GCCTTCAAAC	CACGGCCTAC	GTGACTTCTA	AACAGATAAG	3840
CCCTGACTTA	CAACCATGCC	CTAACTTGCA	TTCTTGCTCA	AAAAGATTAA	ACAAAAGTTT	3900
AAGTTCAGAA	CCCAAAAGCA	ATGACTTTAG	AATTATGTAA	TCAGGTATCC	CTGAGATATT	3960
AAAACACATA	AGAATATTCC	AAATGGGAGC	AAAAGGTTTG	AATACATGAA	AATCAAAC TC	4020
ATATCAGCAG	AGACCATATA	AAGGGCTCTC	ACTGCAGGCT	GACTAGTTAG	GAGGATGGCA	4080
AGGTGATCCA	GGACCTGCGC	ATGCTTTGTC	AGTTCAAATT	GAATCTCATG	CCAACAGCGA	4140
TCTTTTTTAA	CATGTAACAT	TAGGTGTCTC	AGGTACACAT	GACCATAAAC	CACACCTGGA	4200
GGGTTTCTTT	TATTTTCTTT	TTAATATTTT	TCTGAGACAG	GGTCCTACTC	TGTCACCCAG	4260
GCTACCATGC	CCAGCCATGG	AGAGTTTCTT	AAAGATACTG	ATTCCTTTGG	TTAAACCTGC	4320
CACCAAAAAA	AAAAAAAAAA	AAAAAAAAAA	ATACTGATTT	GTGGGCACTC	CATCCCCAAAT	4380
CTATGGAATC	AAAATCTTCT	GGGGGTTTTT	AATAAACATC	TCAAATGAAT	CCTATGATAA	4440
GACAAATTTG	GTAATTGTTA	CACAAACACC	TAATTTAAAA	ATCTGATCAT	TCTACTATCT	4500
AAACACACTC	AGAGTTAATG	AGGGAGAAGG	GAGAAATTGA	TTCTTCTGTA	AGACAGGTAG	4560
CTTTGCAAAA	AGGAAAACAG	CTTAAATCAC	ATTCATTTCT	TATTAAAAGC	TGATGATTAA	4620
TATCATTTTA	GTTTTTCCTG	GGATGGTGAT	ATAATATGGT	GGTCATTCCT	GTCTTAACCA	4680
AAGATATTTT	TGTCCACTCT	AGG TTCACAT	G TAGATTTCA	GCTGGAATTT	TTTTTTTTTT	4740
TTTTTTTTTG	TCCCAGGTAG	ATTCTTAACC	TAAACAAGAA	ATGTAGAAAT	TACAGTTGGT	4800
CCTTGGTATA	TGCAGGAGAT	TGGCTCCACA	ACCTCCCTCC	CCCAGTATAC	CAAAATCCTT	4860
GCATACTCAC	ATCCCACAGA	TTTATTGTCA	GCAAAAGAGA	TGAGAGTTAG	TTTGAACAGT	4920
CTGCCAACAA	TATGATTTGA	TGAATTCTAG	GAAGGTATTT	TCTGCAGTAA	AATATTTCTC	4980
CAACTATCCT	TTTGCCAGTA	TCTAAAATTT	CAGATTAGAG	ATAACTTCCT	ATTCACTAGA	5040
AAAAC TGGAT	TAAAACCTGA	TTAATTAGGC	TTTATTGAAT	ATTAAGGGTT	AAGTATATAA	5100
CTGTGGAACT	TGTAACAGTA	TCACATTTCA	AATTTCTCTT	AAAAC TATAT	CCAATAGAGG	5160

AATGTAAACT	ATTGTCTCCA	CTCAACGAAG	TCAAAGAGTC	CAAAGAGTCT	CCCTGCAGAG	5220
TGAAACATAA	AATAAGCAAA	ATTTCATAGG	CTGCCTGCAC	TACGGCTATG	TGAGGGTTTT	5280
GGTTACCAGG	TGACTGGGAG	TTTCCAAGAA	GGATGCTGGG	AGCCCCATGC	TCTTCCCTGG	5340
GAAACTTTGC	CTTTTCACTA	CTCTACCATC	CAGAAGCAAT	TTTTTAAATG	GGTTATTTAT	5400
TAATTTTCGT	ATTTACACAA	CTCCTACTGA	GATTACTTAA	CATATTTGGT	GGTGACAAGT	5460
TAACAATAAA	TAAGTAAATT	TAAGAATCCT	TGTCCTATAC	CCAACCCAGA	CAATAGAGTT	5520
CTTCCAGACT	CTCCAGCACC	CCCTAGTGGC	ACATATGGAC	CATGGGACGG	GTAGGTAATT	5580
AGCATATATT	TTTCGTTCTG	TTTCCAGCAA	CGGGAAGCAC	TTGGCAAGCA	TCACCTTCTT	5640
TTCTTCGCAA	TACTGCTAGG	AAGTATGTAT	TATGATTATC	TTTATTTACA	TATTAAGAAG	5700
AAACAGTTTT	CAGATAAAGA	ATTTGCTCAG	GGGAACATAG	GTGGCGGGAG	AAAAAAAACG	5760
AGGGTTTACA	ATTTCGGAGC	TCTCACACTT	AATAACCTTG	CTGAAGTATT	GATAGAGGAA	5820
AACATGATCT	TCTTTCAGCC	GCTAACCTTC	TCTGTTTCCT	TTATTGTTCC	TAATACCTTG	5880
TATTACCGTG	GGAGTTACCA	TGTACATTTT	TTTTCTGTG	GGTTTTCTTT	TAATATTTGG	5940
ATTTGGATCT	CCTCCTTTTC	CAGATGTATA	TGTTTAGTTA	TTTAAATTTT	CATGTAATAC	6000
TCTCTAGACA	TATCTCAATC	TTGGTTTTCT	TCCTCTAAGT	TCAATCTGAA	ATATCACTTT	6060
CTCTCTTAAA	TTTGGCTCCC	CCAAGATCCA	ACATTCCAAA	CATATTGCCA	ATGAGTGTAT	6120
ACCTTTTAGC	TTGAAAGCAG	CAGAAAAAAA	GTGGTAAATA	CCTGAGCCAG	GGAACCTAAT	6180
TAGGGGGTTC	TATCAGTGAT	CAAGGCCAGT	GATCAAGGGA	GACACCAGCC	TAATGAAAGA	6240
TGACAGAAGA	TAGCAATACT	CTAATAGAGA	TGTGGTTCAC	AAAGTTCATT	GTGCAGAAGC	6300
AGCTAGGGAG	AGCTTCTAAA	ATACAGAAAT	CTGAGCCCGT	CTTTTTTCTT	TTCTTTTTTT	6360
TTTTTTTTTT	TTTGAGATGG	AGTCTCGCTC	TGTTGCCCAG	GCTGGAGGGC	AGTAGCGCAA	6420
TCTCAGCTCA	CTGCAACCTC	CATCCCCCGG	TTCAAGCGA	TTTAGCTGGG	ATTACAGCCT	6480
TGTGCCACCA	CACATCTGGA	CCCATCTTCT	AATGCAACTG	GTCCACTGAC	TGGCATTTGG	6540
GAATTGCAAT	TTTGCCTCTA	ATTGTAGGAC	AAGGAAGTAA	GAAGAGTTTT	AATCATATTC	6600
AATTCAAGTA	ATGGAGCAGA	TAGATGTAAG	GTCCATCCGA	AAGAGTGAAA	TGATAGAATC	6660
ACAGAATATT	CTTAAAGAAA	GGCAATTTTA	TTCTTTCTAA	CTGCTTATGG	TAAC TACCCA	6720
TGAAAGCAAA	AATATTGATT	GGTAAGGGTC	AATATAATGA	TGTTTCACGA	AGAAAAAGTT	6780

TAATTTGTAA	GTTTTTGTAA	TTCACATTTA	TAATAAATAA	ATCTGTTTCT	GCTTTATAAA	6840
TTTCCTCACT	TGAGTAGATT	AAATATTACC	CTTATAATCT	TCTTTAAACT	TACTGTTTAC	6900
AACCTTTTTA	TTGTCATGAA	GTCAAACATA	AACTTCAATT	CAGCTCGTGA	TCAAAAGATC	6960
ATAAATTCTA	AATAAGTGCT	ATCTGAATTA	ACTTGGTTTG	CTAGAGTTTT	CTGACATTCT	7020
GAAAATTCTA	TATTAGAAGA	ATTCTTTATT	ATATGATAAT	TTATGTTAAA	CAAATTATAG	7080
CAAATTCTAC	ACATAAGGAA	ATTCAGACTA	TATTTATGCT	TAATTATCCA	GGCAGTAGTA	7140
GTACTIONAAGT	AAATATGTGA	GTAAATTTA	TCTGTTTTGA	AACTGTGCC	TCTGTCCTCC	7200
TCTTGATTGA	CAATAAACCC	TCTGTCTCCA	CTTTCACATC	TCCAAAGTTC	AAGTGCATTT	7260
TAATACAATA	TAACAATAAG	CACCATAAAG	ATATAAACTA	TGTTTGTACT	GTTAGCATCT	7320
TATCCCTAAA	TCCAAGCTCA	GGCCCTGGTC	AGTTCAAGCA	TTTGATACAT	ACTTGTCTAT	7380
TAAATCAACA	TTAATCATCT	CTTCATAACT	AGGAAACTA	GGCCAATTTT	ACCCAGATTT	7440
GTCTAAATAC	ACAGATGCCT	ACTTCAGCAA	ACTAAATGTA	GAAGGAAGCA	CATATGAAGA	7500
CAAGGGGGTC	TTTTTTAGCT	GCTATTTACC	AATTAACCCA	ACAATAAAAG	TTTATCACTT	7560
GGCTGGGCGC	GGTGGCTCAC	GCCTGTAATC	CCAGCACTTT	GGGAGGCTGA	GGCAGGTGGA	7620
TCACCTGAGG	TCGGGAGTTC	GAGACCAGCC	TGACCAAAAT	GGAGAAACCC	CAACTCTACT	7680
AAAAATACAA	AATTAGCCGG	GCATGGTGGC	GCATACCTGT	AATCCCAGCT	ACTCAGGAGG	7740
CTGAGGCAGG	AGAATCACTT	GAACCCAAGA	CGGGGAGGTT	GCAGTGAGCC	GAGATCATGA	7800
CATTGCACTC	CAGCCTGGGC	AACAAGAGCA	AAATTCTGTC	TCAAAAAAAAA	AAAAGGGATT	7860
ATCACTTGAT	CTTCAGAAAA	ATAGTGAGGT	CATTATTGTT	TGCTGACAGA	CTACACAAGT	7920
AAAATCTCCC	AAAGGCCAGT	TTTGCCCTGG	CCCTAAGATT	ACTGTAGGGC	CTCAGACATC	7980
AAATCAGTTC	TTCTCATCAC	TCAAAATTCC	CTTAAAATTG	ACCTGACAGA	GAAGCCAACC	8040
ACATTTTTTAA	GCCAAATTGT	TGGGTCTTTT	<u>AAAA</u> ACTAGC	ATTTTGGCTG	TAGTATAACA	8100
GTCTTAGTTT	AACTGATTCA	AACTATGGC	TGGCTTAGTA	AATTTAACGC	TAGTGGCCAA	8160
TAATAACAGA	AAAGAGATAA	ATATTCTTAA	GTATGTATTT	TGAGCCAGGG	ATTCTGCTAA	8220
GTACTIONTATT	CACTCTCATT	AAAGCCTTGA	AACAATTGTT	GCATGTTTAA	GTTATTAATG	8280
AGCCCCATTT	TACAGAGGAA	AATGAGGAAA	CTGACCTATG	TAACTTGCTC	ATGGTCACAA	8340
GCCATTAAAG	GTGGCAGAAT	TAGGATATCA	ATCCAGTCGG	TGTGACTCCA	GAACCCTCCT	8400

ATTTACTCTA TACTACTCAT AAAATTATTT GGTCTTGGGG CTGGGCGCAG TGGCTCATAC	8460
CTGTAATCCC AGCACTTTAG GGGGTTGAGG TGGGTGGATT GCTTGAGCTC AGGAGTTTTA	8520
GACCTGCCTG GGCAACATGG TAAGACCTCA TCTCTACAAA AAAAAAAAAA AAAAAAATAC	8580
AAAAAATTAG CCGGTGTAGT GGCACGCGCC TGTAGTCCCA GCTACTTGGG AGGTTGAGGT	8640
GGGAAGATCA CCTGAGCCCA GGAGGTTGAT GCTGCCGTGA GCCATAATCA TGTCACTGCA	8700
TTCCAGCTTG GGCCACAGAG TGAGACCCTG TCTCAAAAAT AATAATAATA ATCTGGTCTT	8760
GAGAAAAAAT AGTATTTTTT TCTTCATAAA ATATTTTCCA TTTTGAGAAC TTGATTAAGA	8820
AACTCATTGT CTTGCCAATG ACATTACATT CAATCATGCT GAAACATCCA GAAATAGTTT	8880
ACACATCAGT TTGACATCAG TATTATGCAA TTTGAAGCCA CTGTTTGAAA ATAAAAACAC	8940
TGTACCGTGA TTTGTTTATC CAGAGTTCAG ATTATTATAT CCTTGTATAT GAGACAGAAA	9000
CCCCCTTGTA TTCTAGTGCA AACTCTCTTT GGATCTTAAT ATGTATAGTT AACAATAATA	9060
CCATACTACA TTCTAACTAC CTAGAAAGCT AGCATACCTT AACCTGATTA ACTTTTACCA	9120
AGTTACTTGA AATTATAGCA AAGTTACCAT TTAAATCTTG ATTCTGGCCA GGTGCAGTGG	9180
ATGAACCAAG CATGGTGGTG CCTGTAATCC CAGCACTTTG GGAGGCTGAG GCGGGTGGAT	9240
CACGAGGTCA GGAGATTGAG ACCATTCTGG TTAACACAGT GAAACCATCT CTAATAAAAA	9300
ATACACACAA AAAAAATTAG CCGGGCGTGG TGGCAGGCAC CTGTAGTCCC AGCTACTCAG	9360
GAGGCTGAGG CAGGAGAATG CCGTGACCT GGGAGGCGGA CTTGCAGTGA GCCAAGATCA	9420
CGCCACTGCA CTTCAGCCTG GGTGACAGAG CGAGACTCTG TCTCAAAAAA AAAAAAAGA	9480
TTTGATTCTA TCAGTCTACT CACCTTTATA GCTTGACAAT GATTGATTTG TGTAAGGA	9540
TTCAAATCAA AATTTGCAAA CTCCCTTCCT CCAAAGGTAC TCATTTTATA ATACTGAAAT	9600
TCTCTATTAT GTTCTCTGCC CAGTGTCCCA GGGTTTATTG GTTTCTAAAG AGGTAGTGGG	9660
TATATACAGC CTCCCCAAGG GGAATTTAGG AAGTAAGCTG GTTGTCACAA AGACTGGCAT	9720
TAAATAGGTA GAGACCTAGG ATGCTAATAT CTTGCAATGT GCCAAAATAA TTGTCCCTGT	9780
CCCCAACCTC ACCATTGCCA ATATTACCCC TACCCCTCAC AGTGAGCGTC ACAGGCAGGC	9840
AACAACTGG TGTCGTCACA GAATGATTGA TGGAACACAT AGACTGCATT CATTACCTAA	9900
ACATTGTCGT CACACTGCAG CAACCAAAGA CAATCGCATT ACCCAGGGGT TAGATGTAGG	9960
AAGAGTAAAA AACAAAAAAT TTTTGAATGC GTAATTATCA CTAATTATTT TATTTGATCC	10020

TTCAGGAGAA	TGTGGAAGAT	GGCCGAGCAA	GTATCTATCA	ATCTGTCGTT	ACCAACACCA	10080
GCAAAGAAAT	GTCCTGTTTC	AGTGACTTTC	CAATGCCTGA	AGATTTTCCA	AACTTCCTGC	10140
ATAATTCTAA	ACTTCTGGAA	TATTTTCAGGA	TTTTTGCTAA	AAAATTTGAT	CTGCTAAAAT	10200
ATATTCAGTT	CCAGGTATTG	TATTTTTGGG	GAAATGGGTT	TCTCTGCATT	AGTTCAGCTC	10260
ATATTTAGAT	AGAAAAGTTA	CTCTGATAAT	GAAAGCAATT	ATGAATGAAG	TATCCCATTTC	10320
TAAGTATTTG	TTGAAATATA	ACAGCCTCAT	ATAAAACCCA	AAAAGTAGTG	TCATTACCCT	10380
TGGTATTATA	GATTATATAC	ATTAATTGAA	GAGGAAAATC	ATCTGTAA	ATTAAAGGTT	10440
TGAATAATAA	TATATTGATG	TCAAAACTTT	TTTTTTTTTT	TTTCTCCCTG	AGACAGAGTC	10500
TCACTCTGTT	GCTCAGGCTG	GAGTGCAGTG	GCATGATCTC	AGCTCACTGC	AACCTCTGCC	10560
TTCCAGGCTC	AAGTGATTCT	CCTGCCTCGG	CCTCCAGAGT	AGCTGGGATT	ACAGGCACAC	10620
ATCACCATGC	CTGGTTCATT	TTTGTATTTT	TGGTAGGGAC	GAGGTTTCAC	CATTTGGGCC	10680
AGGCTGGTCT	CGAACTCCTG	ACCTCAAGTG	ATCCACCCGC	CTCGGCCCCC	CAAAGTGCTG	10740
GGATTACAGG	TGTGAACCAC	CACACCCAGC	CTCAAAAATT	CATTTAAACT	AATATCTGTT	10800
ATCATTGAAT	ACACCTAGCT	TCATTTGCCT	TGAAAGGGCG	TATACCAAAA	TTAAATTGCT	10860
GTTTTGTTTT	CTTAGCTTCT	TCATAGAAAT	GGGATTTCTT	AGATGTGTAT	TAAATAAATT	10920
CATTGGTCTC	TGTTCACT	AGAAGGCTGT	GGGAAGTATT	TGCTTATCAT	TTTTTTCTGA	10980
ATGCAATCTC	TTACAACCTA	AAGATGGCCA	GATCATTTTG	AAAAACACTT	GGAATTACCT	11040
TTTCCTGTGC	TTCCTCAAAA	TCAACAAAAA	GCAATATTTT	AATTAAGCAT	GCTGAATTTT	11100
TATCAATGGT	CTATACTTTG	AGAAATAGCT	ACTATGCTTA	GAAAATAAAA	TATAAATCAC	11160
ATTTCTTGGC	CAGGTATGGT	GATTCATGTT	TGTAATCCCA	GCAC'TTGGG	AGGCTGAGGC	11220
AGGAAGATCA	CTTGAACCCA	AGAGTCTGAG	ACCAACCTGG	GCAATACAGT	GAAAATCTGT	11280
CTCTACAAAA	AATTTTTTAA	AGATTATCCA	GGCATGTTGA	TACCCACCTG	TGGTCCCAGC	11340
TATTCTAGAC	TGAGAAGGGA	GGATCGCTTG	AGCCTGGGAG	GTCAAAGCTG	CAATAAGTGG	11400
TGATTGTGCC	ACTGCACTCC	AGCCTGGGCA	ACAGTGTGAG	ACCCTGTCTC	AAAGTAAATA	11460
ACTAACATTT	CTGGATAAAT	AACTGTTAGT	GAGGCTTATT	TTTAATACAT	GTCATTTTCT	11520
TAGTAATTCT	AATACTAGGC	TTATATAATA	TCAACTTACA	ATAGTAAATT	TTGGTGAAAA	11580
TTTGTATTTA	TAAATTCCAT	TAAAATGTCC	AGTTCTACCT	AATGTAGTTT	TTCACCAATT	11640

CCTGGTAGAT	CTAACTTGTG	AATAACAGAT	TATGTATACC	AGAAGGTTTT	GTAACTTTGT	11700
GCACTTAACT	ATCAATCTAC	TTAACAAATA	TATTGCCTTT	TTATGATATA	TAACCTTCTAT	11760
TCCATTCTTT	TAAAGATCAT	GTTAGAGTCG	CAAGGAAGTC	ATTTCTCTTG	GTTATTGTGT	11820
TACTGCTACT	TTTGTTCCTT	GGAGAGTGAA	GAGGGGTTGG	GAAGAAAGGT	TTCTGTTTAT	11880
TGGTCTCTGA	GTTGGTGTA	GTCATAGGTG	TTAGAGCTCA	ACTCGAGAAG	CAGGCAAAC	11940
GTAACAAGCC	CTGTTGCTTA	TGATTGTCAA	TGTAATCTAC	ATCAGTGCTT	CTCAAACCTT	12000
AATGTGGACA	TGAATCACCT	GGATATCTTG	TTAAAAATGT	AGGTTCTAAT	TTAATAGGTA	12060
TGGGGTAAGT	TCTGAAATTC	TGCATTTCTG	ACAAGCTTCC	AAGTGATACT	GAAGATCCTG	12120
ATCCTCAAAT	CACATTTTGA	ATAGCAAGGA	TCTACAGCAC	TTAGTTAATA	TACTACTTTG	12180
AACTACCATC	TGAAATCTTT	TCTTTCATCT	GAAAAC TGCC	CAGATATTTA	AAGCCCTTTT	12240
ACAAGATTTT	TACTAATATT	CCATATACAT	TTTTAAATTG	AGACAGCTTA	AAAATTACCA	12300
ACCCAGCAGT	TGGAAAAATA	TCTGAAAATT	TGAGATATAT	AAAAGACTAA	AATACTTGCA	12360
AATGAGAAGC	ATGCCATTCC	TCTAGCATT	TAACTTTG	TTCCACTTGA	CATCGTTTCT	12420
TAATCCAGCA	GATATGAAAC	ATTTATGTAC	AATTTTAAAA	ATTAACAGAC	CTCCAGTGAG	12480
CTACATTTAA	AAAAATCAAT	GAACCAATA	ATCATTTTAT	TCAAATAAGA	TCATGAACTG	12540
TCTTGCTCAC	ATGATGTACT	CTGTTTTTAA	AATAGCAAAT	GTTAAAAACT	ATCATTCAGT	12600
GGAATGCTGA	CCATGTGTCA	GGCACTCTGC	AAAGTGTTTT	GCGTGAAATA	TCTTCTCTAA	12660
TACAAAGTCC	ACAAAGAGGC	GGCTACATA	AACGTTCTTG	ACATATGCCA	ATTGCATGAT	12720
CACTTGAATT	ATTGGTTTGT	TTCCTTGTT	AGATTATCAA	ATAACAAACA	GAGAGAAGTT	12780
CTTTAAAAGA	AAAGATATAT	ATTTGGTGAT	AGAGCATTGT	AATGAGAATG	TACATGCCAT	12840
GGTAAACTAT	TTGTGTATTC	AGGGAGTTAA	AGGAAGACAA	AGGTTTTTAA	ATGGGGAAAA	12900
AATACAATTA	CATAATTGTT	TTGAAATAAT	TATATAAAGA	GCAATAACAA	GGGTGATGCC	12960
AGTCTGAGAT	TGGACAGTTA	CTGAGCAGAT	GTTCTTGTA	AAGTCATTTT	TGTGTAAGAT	13020
TATGATGGTC	TTTGTGTAAG	GTGGTGGTTT	TTGTAGTTTT	TGTTATCAGG	CACACATCAT	13080
GAGAACCCGC	TCTTCTGGC	CTTTCCCAAT	TCTATTTGTC	GGGTTTCTTA	ACATTAGTGA	13140
CTCCATCTAG	ATTCTGACAG	TTTTCATGAG	AACCTGCTTT	TCTTTTCTCT	CTCAAGTCCT	13200
TATTCAGTAT	TCAGCACCT	TAACAGATTA	GTCCCACTGC	TGAGTCAGGC	CTCTTGCATG	13260

AAGCAGCAAT	GAGAAAGACA	CACTTGGCCA	ATGTTATCCT	GGAGTAATTC	TCAATGATGC	13320
CTTCTCTGTG	TTTCTTCAAG	ACAACTGTCC	TTAGTGTGAG	AAAATGTCCA	GATTTCTCAT	13380
CCTCTGGCCA	ATGGAAGGTT	GTCACCTCAGA	GCAACGGCAA	GGAGCAGAGT	GCTGTCTTTG	13440
ACGCAGTTAT	GGTTTGCAGT	GGCCACCACA	TTCTACCTCA	TATCCCACTG	AAGTCATTTT	13500
CAGGTGAGAC	CCGCTGGGAT	TCCCAGCTTT	TTGGAGTAGG	TTTCCAGGTA	CTTTATATGT	13560
AGTTTGGATT	GACAAGCAGG	ATTCATTGCT	GCAACTGGGC	AGAACTTGGC	TCAATAAGAT	13620
TGAGACAGAG	CTAGAAAGAT	GAAAGACACC	AAACATCATC	TTTGTTTCTA	TTGGCCTCTG	13680
AGTCTTCATC	ACACATAGAT	CTCAGAGCCA	ACTTCCTTGG	AAGTCACTAA	GTCCTTGGCA	13740
TAATTTTAGA	GAATTCACAT	CAAACCTGGT	CTCTGTTGGA	GAGGCCCTTT	TAGCCATGTG	13800
CCTGCGTTGG	CCTTTTTTCTA	CCCTGCCAAA	CACCGAGCCT	TTTTACAGG	GCCATACTCA	13860
CACACAAGGG	GAGAGCTCCT	AGAAAGAAAT	GCTTTGCAAG	TTAGTGATGG	GGAGAGAAGT	13920
GCAGGAATAG	AACCCTGCAT	CCAGCTGTTT	TGGTCCACCC	AAGTCTTTCC	TCAGAGAACA	13980
CACTTCTTTC	CCAAGGCCCT	TAGGAAAATA	TGTAATATAG	TGGTTCATAG	TCCAGGCCCTC	14040
ATATTAGAAT	CACCTGGGGA	GCTTCTAAAG	CCCTGATGGC	CTGGAGACCT	ACCCCCAAAG	14100
ATTCAAACAC	TATGGAGTAG	GGTTAGAGCA	ATGAAAGTTT	GCTCAGGTGA	TTTTAATATA	14160
CAGTCAGGAT	TAAGGCCTGC	TCATCTAAAG	CAATTGTTCT	CAAATAGAGT	CACCTGGAGG	14220
GCTTTTGAAA	GCACAAATTG	CTAGGCCCCA	CCCTCCATAT	TTCTGATTCA	ATAGGTGCTA	14280
TGGCTTGAAT	GTCCTGTCCA	AAACTCATAT	TGAGATTAAT	CCCCAATGGG	GCAGTATGAA	14340
GAGGTGGGGC	CTTTAAGAGG	TGATTGAGTA	GTAAGAGCTC	TGCCCTCAAG	AATGGATTAA	14400
GCCATTTGTG	GATAAATAGG	TTAATGGATT	ATTGGGTAC	ACAGGAGTGG	AACTGGTGGC	14460
TTTATAAGAA	GAGGAAGAGA	GACCTGAGCT	AGCATGTTAG	CATGCTTGGC	TCCCTCACCA	14520
TACAATGCCC	TATGCTGCCT	TGGGACTCTT	CAGAGTCCAA	ACCAGCAAGA	AGGCTTTCAG	14580
CAGACGCAGC	CCTTCAACCT	TGACTTCTCA	GCCTCCACAA	TTGTGTGCCA	GAAGAAATAA	14640
CTTCCTTCCC	CTATAAAATA	TTCGGTTTCA	GATATTTTGT	TAAAAACAAT	AGAAGACAAA	14700
TTAAGACAGT	AGCTCTGGCA	TGAGGCTGAG	AATTTGCATT	TCTAACACCA	GGCAATGCTG	14760
ATATTGCTGG	CCATGTGACC	ACACTTTGAG	AACCAATAAT	CTAAAGATTC	TTTCAAGCAA	14820
CCCCACCATC	AATGGCAAAT	ACTTTATAAA	GTCATGTGTT	TCCGTGAAGT	GTAAAAGTAG	14880

TAAC TAGGAA	AGGACACAGA	AGAAGCTTGT	CTGTGATTAA	CCACCAGCAA	GTC ACTGATT	14940
TACACAATAT	GGAAACCAAC	TCCTATGTGC	CTGGT TTTTA	GTTT TAGTTT	TTGTTTACTT	15000
TTTGAAAATA	AGATTGCTAA	ATTGTATTCT	AACTATTACA	CAATTATAAT	AATAGCACTT	15060
CATAATGTGC	TTAAGAAATA	TTTAAGAGTA	TCTGATAAGT	GATTTTTTTT	TTTTTTGAGA	15120
TGGAGTCTCA	CTGTCACCCA	GACTGGAGTG	CAGTGGCACG	ATCTCAGCTC	ACTGCAACCT	15180
CCACAACCTC	CATCTTCCAG	GTTCAAGCAA	TTCTCCTGCC	TCAGCCTCCC	AAGTAGCTGG	15240
GATTACAAGT	GCACGACCAC	CCCTGGCTAA	TTTTTGATTT	TTTAGTAGAG	AGAGCTTCAT	15300
CATGTTGGCC	AGGCTGGTTT	CAAATTCCTG	ACCTCAGTTG	ATCCGCCTGC	CTTGGCCTCC	15360
CAAAGTGCTG	GGATTACAGG	TGTGAGCCAC	CACACCTTGC	CTAATATGTG	ATATTAAAGG	15420
GTCAAATGTC	ATTATATAGT	CCAAAATAGT	ATATAATAGG	CAGGCAGAAG	ACAGTATCTG	15480
GTCCTGCTGT	GTT CATCACC	ATTTATTTGT	CTCTGATAGA	GACAAACTGC	AGCCGTAAGC	15540
TGCAGCCTCT	GAAATAAAAA	ATCAACCCCT	TTGGTCCTGT	TTTTTTGTTT	GTTTTTTGTT	15600
TTGTTTTGGT	GTTGTGACAG	TCTCACTCTG	TCACCCAGAC	TGGAGTGCAG	TGACTCAATC	15660
AGGGGTCACT	GCATTCTTTA	CTTCCCAAGC	TCAAGCAATC	TTCCACCTC	AGTCACCCGA	15720
GTAGCTGGGA	CCACAGGCAT	GCACAACCAT	GCCCAGCTAA	TTTTTGATTT	TTTTGTAGAT	15780
ACAGGGTTTC	ACTATGCTGC	TCAGGCTGGT	CTCAA ACTCC	TGGGCTCAAT	CAACCTGCCT	15840
AGGCCTCCCA	AAGCGCTGGG	ATTACAGGCC	CCACCTGGTC	TGGTACCTAA	ACTTTCTTAT	15900
GTGCTTTACT	CCTATAGAGA	AGAGGCAAAA	CAATTATTAA	CTCCAGAAAG	GAAAAGCTGG	15960
CAATGCAGTT	TTATTGAAAT	TAGCTTGACA	TAGTTGCTCT	GGAGCTCACA	GACTTCTCTC	16020
TTCTTCCCCC	TGAAGGTATG	GAGAGGTTCA	AAGGCCAATA	TTTCCATAGC	CGCCAATACA	16080
AGCATCCAGA	TGGATCTGAG	GGAAAACGCA	TCCTGGTGAT	TGGAATGGGA	AACTCGGGCT	16140
CAGATATTGC	TGTTGAGCTG	AGTAAGAATG	CTGCTCAGGT	GTGATGCTCT	CTGCTTACCA	16200
TGTACCTGGA	GGGGAGGAAG	TGGGGATGCC	ATACTGGAGA	ACCCAGCCA	TATAATCGCG	16260
GCTCCAATCC	TCATTAACTA	GTTGGTTGGT	AGCGCATTGT	GGCATCATAG	AAAATCTGGA	16320
AGTCAAGAAA	CCACTTTACC	TCCTAGCTCT	GTC ACTAACC	AGCCATGAAT	CCTAGAGTGA	16380
TTCATTTTAC	TTCTCTGGGA	GATGGCTCCC	TCATTTTAA	AATGGGAACT	TTTGACCAGA	16440
TGATTTTCCA	TATAAGAGGC	CTTTCATCAA	CATGGCTCAC	TGCAGCCTTG	ACCTCCTGGG	16500

CTCCAATCTT	CCTGTCATCT	CAGCCTCCTG	AGTAGCTGGG	ACTACAGGCA	CATGCCACAC	16560
CACACTCAGC	TAATTTTCAT	ATATTTGTAG	AGATGAGGGT	CTTGCCATGT	TGCCCAGGGT	16620
AGTCTAAAC	TCCTGAACTC	AAGCAATCTG	CCCGCCTCAG	CCTCCCAAAG	TGCTGGGATT	16680
ACAGGCATGC	ACAACCACAC	CCAGCCAAGA	GGCCTTGTTT	CTACCTGGAT	GTTTAATGAG	16740
AGGTAAATCT	GTTTCATATTC	TGGAGGGTGG	CTTTTAGAAA	TTTAGTGTGT	ATTTGAATTA	16800
TATTTGAAAT	ATAGATAACC	TTCAGTTACC	CAAATATTAT	GAAAAGAAAG	ATTAAATAGA	16860
TAGTAGGTCT	CTCAACTAAA	ATCATAGATA	TTTAGGTGCT	TCCTGAGGCC	TTCTAACCAC	16920
TGTCTTCTTT	GCACCTGCTC	AGGAATGACA	CCAGCTGAGC	TGCCAAAGAG	TCAAACATTC	16980
ATTACATGAT	GATGCTGCTG	ACAGTGGTGG	TCAGGAATAG	CAAAACTAA	ACTCCTTCTG	17040
CAAGGACAGA	CCTAGGCAAA	GAAGGGAAAA	TCACTAAACA	TCCTTTCCCA	AAGTATTCCC	17100
TCTCAAGAAG	GCCTGAACCA	GATGCCCAAT	CACTCTTACC	CTAGCTCTTT	CAGCCTGATG	17160
TCTCTGGCCA	CCCAGGGCTT	ACCATGGCCC	TGTGCACAAC	CAACAAATCA	TTTCCATCCT	17220
AAGTCTTACA	CTTTCAGGAC	TCTAGATACC	CAGTGGCAAA	AGTTACAAGC	AAACATGACA	17280
CCCGCCCAGC	AGGTAAATGA	AGGGGTATATA	CTGGGACCTG	TCAGAGTCAT	CTATCAGTCA	17340
GTTAGTTAGT	GCCAGCCCGG	GAACAGAGCA	GGTCACTAAC	ACCGGAAAGA	GACTTACTAG	17400
ACCCAATAAG	TCTTCACTTT	GTGAAAATAA	ACCTCTTGTC	ACTTATCACC	TCAGTGTGAA	17460
GAACAAGTGA	GGAGGCAGGA	ACTGTGACAG	CCTGGAGAAG	AGCAGAGCTG	GAAAATGAGA	17520
GTACCAGCTC	TAGGCTCTTT	CATGCTACGA	ATACCCGCAA	AGCCTTAGGA	ACAGAGTGTA	17580
ATGGGGCAGT	ATGTGAGGAG	CTAATATAGC	AGTCAGCCAA	GTGAAGATCC	ATCCTAGACT	17640
ACTTCACGTT	GTCAGACCAG	TGATTTGGAT	TTAGATCTCT	TCATTCCAAA	GATATCAAAT	17700
CTTAGATGGC	AAGAACCAGT	TCCTTGATATG	GGTCTTGCCC	TACAGGAAGA	CTTATGGTGT	17760
GAGATTCAAT	ATTAAGAAAC	TACCTTGGCT	CTATTTGCAT	GCCTTACAGC	TTCTTAAACA	17820
ATCTTTTGCA	CAGAGTGCAA	AAGACTTTGT	TTCCATCTCC	CTCTATCAGT	GTAAATGCCA	17880
CTAGATGCCC	CCTTTTTTAGG	AGGTACTTCA	CTTTGAGGTC	AATCATCTTT	AAAACAGAGC	17940
CTCAGTAAAT	TCTGGGGCTA	TGCATGTGAT	ACATCACCTA	CATAATAGAT	TCCTCCTAAA	18000
TATAATGTTA	TAATCATACA	TTTCCAGGAT	TATACTCATT	CATCTGCACT	AATCTCTTCA	18060
ATATTTATTA	GAGTAACAAC	ATAAATCTAT	AACTATGATA	AAACCTCTTA	CACAGAGTAA	18120

TATACTCTCA	AGCCTTCTGT	GAAAAGACTA	ACCAGAGACT	TTACAGGAGC	TATACATGCT	18180
AGGAACGGAA	CTAGGCGCAT	CTGCAAACT	TGAAATTACA	ACCTGAACTC	ACCAAAATTC	18240
TGAGTGTGCA	CTGCTCTGTT	AAAAGAAATT	CACCTTCATA	AGGTTACAGC	ACCCTCTACC	18300
ACAATCCAAA	AGCACCCTC	AAGATCATAT	GGGATGGTGC	TGCATCATTG	TATTAGTCCA	18360
TTCTCAACGC	TGCTATGTAG	ACATACCCGA	GACTGGGTAA	TTCATAAAGA	AAAGAGGTTT	18420
AATTGACTTA	CAGTTTGGCA	TGGCTGGGA	AGCCTCAGGA	AACTAACAAT	CATGACGGAA	18480
TGCACCTCTT	TACGAGGCTG	CAGGAGAAAG	AATGAGAGCG	ACTGGGGAAC	CCCTTATAAA	18540
ACCATCAGAT	CTCGTGAGAA	CTTACTCCCT	ATTAGGAGAA	CAGCATGGCA	GAAACCTCCC	18600
CCATGATTGA	ATTATCTCCA	CCTGGTCCTG	CCCTTGACAC	GTGGGGATTA	TTATAATTTA	18660
AGGTGAGATG	TGGGTAGGGA	CACACAGCCA	AACCATATTA	GTCATTTACA	TACTTCTGAC	18720
CAAAAACCAA	ATCTCTGGCC	TTTGACCTAA	AACATGCGTC	TCAGAGAAAG	CAGCCTGAGC	18780
CTAAATCCTC	ATGTTTCTCT	CACTGTTGCA	GCTAGTGTCA	TTAAGGCAGG	TTAGACCACC	18840
CTGCTGTAGG	GAGGGTCACA	ACAGAAAAAG	AGTGAATCAA	ACGGGCAGAG	CATACCATTT	18900
GAAACATGGT	TTGCTCCTGA	GAAAGAAGAG	GGGACAGTAA	GTAATGGAAA	GAGACACTAA	18960
TGAAAATATT	TTTGTATCTA	ATATCTAATC	AAAGTATTGC	CAAGTCAGCC	TATAAGGGCA	19020
ACGGCAGGAG	AAATTCAGAA	CATAGGTATA	TACCACACAC	AGACCAGCAA	TATAGGAATG	19080
CTTGGTATAG	GTGCTACTTC	ACAAGCTAGG	AATGTAAGGC	CCATCCCCAC	AAAATTTGTC	19140
TCCAAATTCT	GGTTTACTCC	AGACATAAGG	CACTGTATGA	AACTCCTCTC	TTCCAGCCTA	19200
ACTTTATAAC	TTAACAGCTA	GCAGTACTTA	TCACTTGCCA	GGCAATATTT	CAAGTACTTT	19260
ATATATACCA	CCTCATTTAA	TCTACACAAG	AATGCCATGA	GGTAGGTACT	GTTAATACCC	19320
CCATTTTACA	GAGAGAGAAA	CTGAGGCACA	GAGAGATTGA	AATAATTCAA	CCATGGCAAC	19380
ACAGATTGAA	ATAGTTCACC	CACAGTAGTG	TGATTGGGAT	TCAAACCCAA	GCAGTCTGTA	19440
TCCAAACCTC	TCAAGTAAAT	TGGTTACCTT	GCAAGTGAAT	CTTATGTGTT	TATCAAGTAT	19500
AGCCTTAAAC	AAAACTTAT	TGCATGGTAT	GTAAAAATTT	AAGAAGCAGT	TCAAGTATGC	19560
ATTTGGCCAA	TGGGGGAGTA	ACAGCAAACA	CAGCAAAATA	TACATTTGAA	AAGAGATTAA	19620
ATGTACATTT	TGGAAACAAG	GGAAATCTTA	ATAACAAGG	TAAAGAATAC	ACCTGAAAGA	19680
GGATTCAGAT	GTGCACTTGA	AGAGAAAGAG	AATCACAGTA	TAAGTTCAGA	GTTTTTAACT	19740

TTTAAATAC	ATTACAAGCA	CTGTGTCTCA	TGCCTGTAAT	CCCAGCACTT	TGGGAGGCTG	19800
CGGCAGGAGG	ATTGCTTAAG	CCCAGAAATT	TGAGACCGAC	CTGGGCAACA	TAATGAGACC	19860
GTCTCTACAA	AAAAATTGTT	TGAATTAGCT	GGATGTGGTG	GTACATGTCT	GATACTGAGG	19920
TGGGAGGATC	ACTTGAGCCT	GGGAGGTCGA	GACTGCAATG	AGCTATGACT	GCACAACGTC	19980
AGTCCAGCCT	GAGTGACAGA	GCAAGACCCT	GTCTCACACA	CACACACACA	CACACACACA	20040
CACACACACA	CAAAATAAAG	TCTTTTAAGT	ATGGAAGGAA	GATTATTTCC	CCTGTTATTC	20100
TCCATCCAGG	GATATTCAGA	TGCATATACA	CTTATACTTG	TGTAGTCACT	AGGCTATAAT	20160
CGCACATTTT	CAAGGATTAT	AATCATTCTA	CCTGCACTAT	AGAAGAACT	TAGGTGAGTG	20220
GAAAACATGA	GAGGAGGGAG	GGAGGAACTT	TCTCTTAAGG	AGCAGCAAAC	CACAACGTGA	20280
AACATGGGAA	AGACTTGTGG	ATTTTATCAT	CAGAGTTAGC	CCAAAGACTT	TCTCGTGTCT	20340
CCATGAAGTT	CTCAAGATTT	TGTTGCAGTC	TTCCTGCATC	AGTGTAATG	CCACTGGGTA	20400
CCCCTATTTA	GGAGGTACTT	TACATTGAGG	TCAATCATCT	TTAAAACAGA	ACCTCTGTAA	20460
ATTCTGGGGC	TACACATGTG	ATACATGACC	TTCATAGTAG	ATTCCTCCTA	AACGGGACAA	20520
TGCCCTAATT	TAAACTGCAT	TTCTTTTTGC	TTGCCAGGTT	TTTATCAGCA	CCAGGCATGG	20580
CACCTGGGTC	ATGAGCCGTA	TCTCTGAAGA	TGGCTATCCT	TGGGACTCAG	TGTTCCACAC	20640
CCGGTTTCGT	TCTATGCTCC	GCAATGTACT	GCCACGAACA	GCTGTAAAAT	GGATGATAGA	20700
ACAACAGATG	AATCGGTGGT	TCAACCATGA	AAATTATGGC	CTTGAGCCTC	AAAACAAGTA	20760
GAGTTATTTT	GCTTTTTTAA	TGGTATACTC	GTTGGTGAGC	AAAGTTGTCT	GAAGGTGTCT	20820
CCCTTAACAA	AGATTCAAAT	TGCTAACACG	GTAGTTAAAA	CTACAATCTA	ACAATATGAG	20880
TATCTTATAG	GTCCTGGAGT	TTAGCTTCTA	AATTTGGTCT	GTATGCCTTT	AAAAAATACT	20940
TAAGAAGATG	AAGCAGAAGT	GTTATAAGCT	GCTCCAGAAA	GCAAAACTAG	GGGAGAACTT	21000
TCTAATACCC	AGAGTTATCT	AACATTGGAG	AAAACGTGTT	CAAGAGATTA	CGACCTGCCT	21060
TTCAGAGGGG	TGTGGTGGGA	AACATGTAAT	TCTCCATCTA	ATAATTTATG	CTTTGCTAAC	21120
CCTATAGCAT	GAAGGTTCTT	CCCATGGGAA	ACCTTTGAAA	ACACATTCCT	TTTTCTTTGC	21180
TAAAAGACAA	ATCTCTGTTG	ACGTCAAAGT	TATATGTCAG	TGATTTAAGC	ACAAGCAAAT	21240
GTTATGAATG	GTTCTTTTGC	TTTAGTTGTT	ACAGGCTTCT	TCCCTTAAAA	AAACAGAAGA	21300
GCTTTAGAAT	CTTTTAACAA	ATGCCTGCCG	TGCAACTACC	ATATTCTAAG	ATCTGACATA	21360

AGTGCCACGT	ATCGTCTATT	AAAAAAGAA	AAAGAAAATG	TTCTCAAATC	TACAAAAAAA	21420
ATAAGCGGAC	TTTGCATCAA	CATCCATGCT	ATTACTAACA	GAGACTCCAT	GGATATTTGG	21480
GATTAACAAA	TATCACCAAA	CCTAATTTTA	TACATTAATT	TTCACATTGA	TCCCTTCATA	21540
GATTTCAAAA	CTAGTGGAAA	TTTAGCAAAT	TTTTTCTTAT	GATCAAATAG	GGGTTAAATA	21600
AAACAGCAAA	ATAATAAAAAG	CTAGATAGCA	TGAAAAAGGT	TAAAAACAGA	AATGGTATAA	21660
TAACCACCAT	AATACTTGGG	GATTGACCAT	AGGCACAGGC	ATTTTGTCTA	AGCCCTTGGG	21720
GATGCTTCCT	TCCTTAAAAT	CTCTTTCAC	CACGTTGCCT	ACATGTTTTT	CCTTATTTAT	21780
TGACAAGAGA	TATTTGTGAC	ATGAGAATTA	AGTCAGAAAA	TAAGGATTTG	CACAGACAAC	21840
CAGTTAAGTT	AGAGTTTTAC	AGATATTTGA	AAAGCCCTTT	TATTTTCAGA	GCCGTACCCC	21900
AAAAATATCA	AGAGGGTTCA	AGATTCCCTCA	GCAAATGATC	CTTCAGAATG	TTTTTCTTCT	21960
GTATGTCTCA	GATACATTAT	GAAGGAACCT	GTACTAAATG	ATGATGTCCC	AAGTCGTCTA	22020
CTCTGTGGAG	CCATCAAGGT	GAAATCTACA	GTGAAAGAGC	TCACAGAAAC	TTCTGCCATC	22080
TTTGAGGATG	GAACAGTGGA	GGAGAACATT	GATGTCATCA	TTTTTGCAAC	AGGATATAGT	22140
TTCTCTTTTC	CCTTCCTTGA	AGATTCACTC	GTTAAAGTAG	AGAATAATAT	GGTCTCACTG	22200
TATAAATACA	TATTCCTCCG	TCACCTGGAC	AAGTCAACCC	TCGCGTGCAT	TGGTCTCATC	22260
CAGCCCCTAG	GTTCCATTTT	CCCAACTGCT	GAACCTCAAG	CTCGTTGGGT	GACAAGAGTT	22320
TTCAAAGGTA	AGTGTGTAGG	CAGGTGAGTG	GCTAAGCGTT	TCAGATCTGG	TGAAGTTTAT	22380
CAATAATGAT	AAGAAGGTTG	CCTGAGATAA	AAAGGTTGCC	AAGAAAAAGT	TTGACAACCT	22440
TGGCTGCTCT	CACAAGACTA	ACATTCTAAA	AAGTTACTGG	AGAATTCAAA	GAATAACAAA	22500
TACAGGAATT	TAGTAATAAT	AAATACCTGC	AATCATCCTT	TTAAAAATAT	AGACAGTCAA	22560
GAGAATTTCA	ACTGGCATAA	AGCTAAGTGC	ATGTTAACTT	TTCTTTGAAT	CGTGAGAGAT	22620
AAGTTTAAGA	AAAAGATCTG	TCTCCTGGTT	TACCTCTGT	GTTGTTTAAA	AATTCCTCAG	22680
CATATCTGCA	AATCAATTTA	ACTCTTAATA	CTTGAGCAGC	TCAACCTCAC	AAATCCCTAC	22740
AAGTTATAAA	ATTATTAAAA	GGTTTCTTTC	TGGGTGTCTG	TGTAGCACTT	CATACTCCTC	22800
AGAACGGTGT	TACCTCCCTG	CCTCCAGGGT	TCAATTCTGT	TCAGCAAAAG	CTTACTGAAT	22860
ACCTTGCCCT	GTGCTGGGAA	CTGGTGGGAC	AGAGAGAAAT	TTAAACAGAT	CATTTCAACA	22920
TAACATGACA	AATGCTTTGA	TTGAATAATA	TATGGAGTGT	TCAGGGAAGG	AGAGAAAGGG	22980

CACTTATCAT	GGTAGAATAA	GGGAAGGGCA	CATGATAAAG	GAAAACGTCC	TGGATAACTG	23040
CATTTCTCAG	GGGCAGAAAA	GGGGATTGCC	TGAACAAAAG	CATAGAGTCA	ATGATGCATA	23100
TGGAAGGGCA	CATGCTATTT	GACATTGCTA	GAGCATGACG	TATGAGGCAG	AGAGAGATGA	23160
GCCATTACTC	TTGGAGAAGA	AGGAGACAGG	ACACAGGAAT	TTTTTAAGAC	ATGCTATGGA	23220
GCTTAGATTA	TAAATTATAG	ATCAGTTCTT	CCCAAATATG	GCTACATATG	AAAATCATCT	23280
GATGGATCCT	TAGGGACCCT	GATTAAGTAA	GACTGGCCAA	GGGACCTGGA	ATCTGCATTT	23340
TAGAAAGCTC	TTCAGCCCGG	GGCACCAATG	AAGGGTTATA	AGCAAGGAAC	AGGCATTAGC	23400
AGATTTACAC	TTCAGATAGA	TTGTTTCAGC	AGTAGTGTGG	AATATAGATT	TGAAAGTGGG	23460
GAAAGACTAC	AGCCTCAGGG	ATGAAAGAGA	AAGCTACTGA	AATAGCCTAT	GCTAAAATAT	23520
GATGCATCCT	GGGCCAGGGC	AGAGATACAA	AGTGGAAAGG	AAGCCATAAA	TGTGAGAAAT	23580
CATTAAGGGA	AAAATCAGCA	TGACATTATG	ATTGGTTCAA	TGTGGGAAAG	TCAGAGAAAT	23640
AGAGAGGAAT	CTAGGAGGAC	TTACAGATCT	CTGGCATTGG	AAACCAGGTG	GACAGTAGTG	23700
CTGTGAATAC	AGAGGGGGTG	TGCAGAAAAT	GATGCAAGTC	TGGACAGGAG	GGCTTCAGTG	23760
AGGAGCTCAG	GTCTGGACTA	CTTGAACATG	AGATGTCTGA	TGACTCTAGG	CAAGGGGACT	23820
TGACCATATT	TCAACACATC	CAAAGCTCAG	GGGACACTTG	TGGGCAGGCG	ATGGAGTCAT	23880
GAGCACACAG	TAATAACTTC	TGCATCAATC	TTTCCCTATC	TCTACTGCCC	TACTCTCATC	23940
TCTCACCAGG	TTTATTTCAA	CAGCCTCTTT	ACTGGTCTCC	CCAGCTTTGG	GCTTGCCTCC	24000
CTGGAGTCCA	TTTTCTTAAA	TTCAGCAGCC	AGACAGATCT	TTCCAAAAAA	TAAATCTGAT	24060
CTTCTCACTT	CATTGAGAAT	ACTCTTCCAC	TGATTTGATT	TGGGGCCTCC	TGTCACCTTC	24120
AGGATAGAGC	CCAAACCACT	AGTCATGGCT	GCCAGGCTCC	CAGACACACT	TCCCTTTTCC	24180
AGCCTCTTCT	CTTGGCCCTC	TCCACTTGTA	GTCCATGCCG	TAGACTGTGC	ACCCTGGACA	24240
GTGTCACATA	GAGTGCTATG	GGGGTGGCAC	CCCCTGAAGT	TCAACAGCAC	GGAAGCCCTG	24300
ACTGGTATGA	CATGGTTCAA	TGTCCAGAGT	TTAATTTTAA	GAATCAACAA	CTAGACAAAG	24360
TAATGATATT	GACTCAAAC	TACTATTCAA	ACCAACCTTT	TATTCCTTAG	GCTTGTGTAG	24420
CCTGCCCTCA	GAGAGAACTA	TGATGATGGA	CATTATCAAA	AGGAATGAAA	AAAGAATTGA	24480
CCTGTAAGAA	TTTTTTTTTAA	TTCTTTACAT	GAAGCAGTGT	TTCTCAAAGT	ACAGTGATCT	24540
AACTACTTAC	AAGAACCACC	TAGCTGCCTG	ATAAAATGCA	AATTTCTGGG	CTATAGCCCA	24600

GATGATTGAA TCAGAAACTC CGTGTGTGAG GCTAAAAAGT TGCATTTTTTA TCTTCTTCCT	24660
AAGCGATTCT TATACATACT AGGTTAAGAA CCAAATACTT AAAGATAAGA ATTGTACCAA	24720
ATCAGAGCAC TTCTCCTTGG CTTAATTTCA TTTCAGTTTT ATATGATGCC TATGTCAGAT	24780
TCCATAACTT CTCAAGCCAC CTACACTCTG TGGTTAGAGA GGAATGGGA TGAGACAGTG	24840
GTGGTGATAG TAGCTTGAAT AGCTGTGAAA AGTTAGAGAA TCCCCATCAG AATAAATTAG	24900
GAAGGGGTTG GTGTGAAGGT TCAAGGATTT GTACTTTGTG ATGAGGTAAA ATGAGGTTCA	24960
ACAGTGATCG AGTACCCTTG GAAAGTTGAT TTGGGGCTTA CATCAGGTGT AAAGAGTTTT	25020
CTCATGTTCA AATTCAAATT TACCTAAGAT TGATTGAGTA TCTACTATAC GCCATCCAGA	25080
CTGCCAGGTA CTTTAGTAAT TTAACAAGCA AATATTAAGC ATCTCCTTTG AGCAAGACAC	25140
CAAGCTATGC TTTCATATGC ATTATCTCAT GAATTCCTGC AGCCGCCCTG GCTAGCATGT	25200
ACTTGCCTGG AGATTTGCCA CCGCTTAAAA AATGCCAAAC AATGGTTACC AATCTTGTC	25260
CATTTCTAGA GCATCCATGA ATTCATGGCT CTTTATTTGA GGGCGTATTC TCAATCTGAG	25320
ATATGAGCCT CCTGGTATGA TAAACTCAAA CTTTCCACCA GAGATTCATT GAAAACTCAT	25380
TCACATATTC ACTCATTCCT TCATTCCTTT AGCAGTTTTG AATGCCTAAT ATTCTAGAAA	25440
ACTTAGAACA TTCTGTGAAC ATTCCCTTTT TACTTTCTTC ACTAAGGTTT GGAGAAAGCC	25500
AGAGCCAGAC GTTGCAGACC AATTATGTTG ACTACTTGGA CGAGCTCGCC TTAGAGATAG	25560
GTGCGAAGCC AGATTTCTGC TCTCTCTTGT TCAAAGATCC TAAACTGGCT GTGAGACTCT	25620
ATTTCGGACC CTGCAACTCC TATTAGTATC GCCTGGTTGG GCCTGGGCAA TGGGAAGGAG	25680
CCAGAAATGC CATCTTCACC CAGAAGCAAA GAATACTGAA GCCACTCAAG ACTCGGGCCC	25740
TGAAGGATTC ATCTAATTTT TCAGTTTCTT TTCTGTTGAA AATCCTGGGC CTTCTTGCTG	25800
TTGTTGTGGC CTTTTTTTGC CAACTTCAAT GGTCCCTAGTC AGCATAATGC TTTGGGCTTT	25860
ATTATCTTGT CAGTCACTAC CTCCTAAAGA AAAAAAAAAA GGCTAGAAGA AAAAAACATTA	25920
CATTCATGTT CTAATTATAG ATTTTAGAGT TAGGTAGTAC AGGTAAGGGG GAAATTGTAA	25980
AGAATTAGCA GAATTAGGCA TATGTACAAA ACCAAA	26016

(2) INFORMATIONS POUR LA SEQ ID NO: 2:

(i) CARACTERISTIQUES DE LA SEQUENCE:

- (A) LONGUEUR: 1731 paires de bases
- (B) TYPE: nucléotide
- (C) NOMBRE DE BRINS: simple
- (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADNc

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 2:

AACCAAGGGA GAAAACTATT CTGTCAAAGA GACGGTGCCA AAAGGCAAAA ACAAAGGAGC	60
TGATGGCAAA GAAGGTAGCT GTGATTGGAG CTGGGGTCAG TGGCCTAATT TCTCTGAAGT	120
GCTGTGTGGA TGAGGGACTT GAGCCCACTT GCTTTGAGAG AACTGAAGAT ATTGGAGGAG	180
TGTGGAGGTT CAAAGAGAAT GTGGAAGATG GCCGAGCAAG TATCTATCAA TCTGTCGTTA	240
CCAACACCAG CAAAGAAATG TCCTGTTTCA GTGACTTTCC AATGCCTGAA GATTTTCCAA	300
ACTTCCTGCA TAATTCTAAA CTTCTGGAAT ATTTTCAGGAT TTTTGCTAAA AAATTTGATC	360
TGCTAAAATA TATTCAGTTC CAGACAACCTG TCCTTAGTGT GAGAAAATGT CCAGATTTCT	420
CATCCTCTGG CCAATGGAAG GTTGTCACTC AGAGCAACGG CAAGGAGCAG AGTGCTGTCT	480
TTGACGCAGT TATGGTTTGC AGTGGCCACC ACATTCTACC TCATATCCCA CTGAAGTCAT	540
TTCCAGGTAT GGAGAGGTTT AAAGGCCAAT ATTTCCATAG CCGCCAATAC AAGCATCCAG	600
ATGGATCTGA GGGAAAACGC ATCCTGGTGA TTGGAATGGG AAACTCGGGC TCAGATATTG	660
CTGTTGAGCT GAGTAAGAAT GCTGCTCAGG TTTTATCAG CACCAGGCAT GGCACCTGGG	720
TCATGAGCCG TATCTCTGAA GATGGCTATC CTTGGGACTC AGTGTTCCAC ACCCGGTTTC	780
GTTCTATGCT CCGCAATGTA CTGCCACGAA CAGCTGTAAA ATGGATGATA GAACAACAGA	840
TGAATCGGTG GTTCAACCAT GAAAATTATG GCCTTGAGCC TCAAAACAAA TACATTATGA	900
AGGAACCTGT ACTAAATGAT GATGTCCCAA GTCGTCTACT CTGTGGAGCC ATCAAGGTGA	960
AATCTACAGT GAAAGAGCTC ACAGAACTT CTGCCATCTT TGAGGATGGA ACAGTGGAGG	1020
AGAACATTGA TGTCATCATT TTTGCAACAG GATATAGTTT CTCTTTTCCC TTCCTTGAAG	1080
ATTCACTCGT TAAAGTAGAG AATAATATGG TCTCACTGTA TAAATACATA TTCCCCGCTC	1140
ACCTGGACAA GTCAACCTC GCGTGCATTG GTCTCATCCA GCCCCTAGGT TCCATTTTCC	1200

```

CAACTGCTGA ACTTCAAGCT CGTTGGGTGA CAAGAGTTTT CAAAGGCTTG TGTAGCCTGC      1260
CCTCAGAGAG AACTATGATG ATGGACATTA TCAAAAGGAA TGAAAAAAGA ATTGACCTGT      1320
TTGGAGAAAG CCAGAGCCAG ACGTTGCAGA CCAATTATGT TGA TACTACTTG GACGAGCTCG      1380
CCTTAGAGAT AGGTGCGAAG CCAGATTTCT GCTCTCTCTT GTTCAAAGAT CCTAAACTGG      1440
CTGTGAGACT CTATTTTCGGA CCCTGCAACT CCTATNAGTA TCGCCTGGTT GGGCCTGGGC      1500
AATGGGAAGG AGCCAGAAAT GCCATCTTCA CCCAGAAGCA AAGAATACTG AAGCCACTCA      1560
AGACTCGGGC CCTGAAGGAT TCATCTAATT TCTCAGTTTC TTTTCTGTTG AAAATCCTGG      1620
GCCTTCTTGC TGTGTTGTGTG GCCTTTTTTTT GCCAACTTCA ATGGTCCTAG TCAGCATAAT      1680
GCTTTGGGCT TTATTATCTT GTCAGTCACT ACCTCCTAAA GAAAAAAAAA A              1731

```

(2) INFORMATIONS POUR LA SEQ ID NO: 3:

(i) CARACTERISTIQUES DE LA SEQUENCE:

- (A) LONGUEUR: 535 acides aminés
- (B) TYPE: acide aminé
- (C) NOMBRE DE BRINS: simple
- (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: peptide

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 3:

```

Met Ala Lys Lys Val Ala Val Ile Gly Ala Gly Val Ser Gly Leu Ile
1              5              10              15

Ser Leu Lys Cys Cys Val Asp Glu Gly Leu Glu Pro Thr Cys Phe Glu
              20              25              30

Arg Thr Glu Asp Ile Gly Gly Val Trp Arg Phe Lys Glu Asn Val Glu
              35              40              45

Asp Gly Arg Ala Ser Ile Tyr Gln Ser Val Val Thr Asn Thr Ser Lys
50              55              60

Glu Met Ser Cys Phe Ser Asp Phe Pro Met Pro Glu Asp Phe Pro Asn
65              70              75              80

Phe Leu His Asn Ser Lys Leu Leu Glu Tyr Phe Arg Ile Phe Ala Lys
              85              90              95

Lys Phe Asp Leu Leu Lys Tyr Ile Gln Phe Gln Thr Thr Val Leu Ser
100              105              110

```

Val	Arg	Lys	Cys	Pro	Asp	Phe	Ser	Ser	Ser	Gly	Gln	Trp	Lys	Val	Val	115	120	125
Thr	Gln	Ser	Asn	Gly	Lys	Glu	Gln	Ser	Ala	Val	Phe	Asp	Ala	Val	Met	130	135	140
Val	Cys	Ser	Gly	His	His	Ile	Leu	Pro	His	Ile	Pro	Leu	Lys	Ser	Phe	145	150	155
Pro	Gly	Met	Glu	Arg	Phe	Lys	Gly	Gln	Tyr	Phe	His	Ser	Arg	Gln	Tyr	165	170	175
Lys	His	Pro	Asp	Gly	Ser	Glu	Gly	Lys	Arg	Ile	Leu	Val	Ile	Gly	Met	180	185	190
Gly	Asn	Ser	Gly	Ser	Asp	Ile	Ala	Val	Glu	Leu	Ser	Lys	Asn	Ala	Ala	195	200	205
Gln	Val	Phe	Ile	Ser	Thr	Arg	His	Gly	Thr	Trp	Val	Met	Ser	Arg	Ile	210	215	220
Ser	Glu	Asp	Gly	Tyr	Pro	Trp	Asp	Ser	Val	Phe	His	Thr	Arg	Phe	Arg	225	230	235
Ser	Met	Leu	Arg	Asn	Val	Leu	Pro	Arg	Thr	Ala	Val	Lys	Trp	Met	Ile	245	250	255
Glu	Gln	Gln	Met	Asn	Arg	Trp	Phe	Asn	His	Glu	Asn	Tyr	Gly	Leu	Glu	260	265	270
Pro	Gln	Asn	Lys	Tyr	Ile	Met	Lys	Glu	Pro	Val	Leu	Asn	Asp	Asp	Val	275	280	285
Pro	Ser	Arg	Leu	Leu	Cys	Gly	Ala	Ile	Lys	Val	Lys	Ser	Thr	Val	Lys	290	295	300
Glu	Leu	Thr	Glu	Thr	Ser	Ala	Ile	Phe	Glu	Asp	Gly	Thr	Val	Glu	Glu	305	310	315
Asn	Ile	Asp	Val	Ile	Ile	Phe	Ala	Thr	Gly	Tyr	Ser	Phe	Ser	Phe	Pro	325	330	335
Phe	Leu	Glu	Asp	Ser	Leu	Val	Lys	Val	Glu	Asn	Asn	Met	Val	Ser	Leu	340	345	350
Tyr	Lys	Tyr	Ile	Phe	Pro	Ala	His	Leu	Asp	Lys	Ser	Thr	Leu	Ala	Cys	355	360	365
Ile	Gly	Leu	Ile	Gln	Pro	Leu	Gly	Ser	Ile	Phe	Pro	Thr	Ala	Glu	Leu	370	375	380
Gln	Ala	Arg	Trp	Val	Thr	Arg	Val	Phe	Lys	Gly	Leu	Cys	Ser	Leu	Pro	385	390	395

Ser Glu Arg Thr Met Met Met Asp Ile Ile Lys Arg Asn Glu Lys Arg
405 410 415

Ile Asp Leu Phe Gly Glu Ser Gln Ser Gln Thr Leu Gln Thr Asn Tyr
420 425 430

Val Asp Tyr Leu Asp Glu Leu Ala Leu Glu Ile Gly Ala Lys Pro Asp
435 440 445

Phe Cys Ser Leu Leu Phe Lys Asp Pro Lys Leu Ala Val Arg Leu Tyr
450 455 460

Phe Gly Pro Cys Asn Ser Tyr Xaa Tyr Arg Leu Val Gly Pro Gly Gln
465 470 475 480

Trp Glu Gly Phe Arg Asn Ala Ile Phe Thr Gln Lys Gln Arg Ile Leu
485 490 495

Lys Pro Leu Lys Thr Arg Ala Leu Lys Asp Ser Ser Asn Phe Ser Val
500 505 510

Ser Phe Leu Leu Lys Ile Leu Gly Leu Leu Ala Val Val Val Ala Phe
515 520 525

Phe Cys Gln Leu Gln Trp Ser
530 535

(2) INFORMATIONS POUR LA SEQ ID NO: 4:

(i) CARACTERISTIQUES DE LA SEQUENCE:

- (A) LONGUEUR: 25464 paires de bases
- (B) TYPE: nucléotide
- (C) NOMBRE DE BRINS: double
- (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADN (génomique)

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 4:

TCCAGTCCTG CAGCAACCTC CTAGTTCCTG CTCTTTCAGC TCTTTGACCT TTTGCAAGCA 60

CCTAATTCCC TGTAGTATAT ACCTTTCTTC ATGATATATA GTGTTTTTTA TCTCCTGCAC 120

TAAATCATGA GCATATGCAT ATAAATCATA ATATGAAATC TTAAAAACAG AAGTACTTTT 180

GCTGAGGCAT TAAGCATATA ATCAGTCAGC AGGTCCCCAA ACATCTAATT CCTGAATATC 240

TCATATATCC TGTCTCCATT ATCCATTCCCT CTAATGCTAC TCTAATTTAA GTCCTCAGTC 300

TCTCTGGCCT AGATTGTTGA AATAACATCC TGGGTTTTTG GTCTCCTTGA TTCTAGTCAC 360

CATCCTCTCT AGCCTCCAGG TGAATCTGAT CTTGTCTGAT GTTGTCACCT CTTGTTCAA 420

AATTCTCGAA	TGGACAACCG	TAATCCAGAA	GGTAGTATCC	AAACCTGTGA	TTGTGGCACT	480
TCAGTATCCT	TCATAACCTA	TGTCCTGCAT	GTTTAACCCA	TATTTTGCTA	TTCCCATCAC	540
TTATGGTCCA	GCAAAACTGA	ACTAATTGTA	GTTCCCCCAT	CACGTGTTCT	TACTTTTCTA	600
TGCATTTTCA	CATATTTTTC	TCTCTGCCTT	TTTTCTATTT	CTTGTCCCTT	ATCTGTCTGG	660
AAAACATCTA	TTCTTCCTTC	AAGACTCAGC	TGTCTTCTCA	CACTCCTTGA	AGCCTCTCTT	720
TCCTCCTCCA	AGTGGACCTA	GATTTTTTCTT	CCTACATGCT	AGCACTACAC	TGAACCATAC	780
TTCCACTGTG	ACATTTATCA	TCTCCCTCAA	CACTAGACTT	CATGGTTCCA	GATGGAAAGC	840
ACTGTGTCTT	CTCACCTTTG	AATCCCCCAA	AAGACTATTA	TAATGCATGA	CATATAGTAG	900
GCTGTCAGTA	CAGTGAAAGG	AATGGCCAGA	GGAAGGAAAG	GAGGGAAACA	GAAGCAGAAA	960
GGACAGGTAT	AGAAGCCGGA	GGGAGCCAGA	GACAAGGTTT	AGAGACCACA	ATTCTGTCTT	1020
TTGAGTTCAC	TAGTTTTACA	AGCTCATCTA	TAAGCGTTAG	TTCAGCAACT	CAGATCAGGC	1080
CCTAAGTTTC	CAGAAATTTG	AGCTACTTTT	CACTGTTGGC	ACAACAAAAC	GTTTCATTAT	1140
AGTCCAGGTG	CATAGCCTTT	GTTTATATAT	TCTATATTTT	CAAAGCAAAC	ATAAATGAAA	1200
GAATCATGTG	TCCCCTAATC	TCCCAGGAGT	TTCACCTTAC	AGCTCCAGTG	GCCATGGCAG	1260
TCAGTGTTTT	ATATTTTTTG	TAACAAGAAC	CAAAGACTTC	ATTCTTCCTT	TTTCCTACCC	1320
CTTTCTTTTT	ACTTCACCCA	TGCCTCCCCT	GTTCTTCTCT	TATCCCTACC	ACACTCGTCC	1380
TTCTCTTTCA	GATTTTACTA	TGGCTCTATA	CCATTAAAAA	TACAAGAAAA	AAAAGGAATT	1440
TTACTTTAAG	AATAACTCCT	CCCCCTTCCC	CAGTTTTTCA	ATCAAAAGAC	ATTGTTAAAT	1500
GCCATTCTCT	TCCACATTTT	GAGAACTGCT	GATTCTCTGG	GGAGAGAAAG	GTGATTGCTT	1560
AAGAGGTGAA	GTCCCTTAGA	GCATTCAAAA	TGAGGAGTGA	TTCTGTACAG	AGGATATCAT	1620
GCAGCAGGCT	GGATGTCTAG	TTCCAATTCC	TTTATTTGTT	ACCTCTGAGA	CCTTGAAGAA	1680
GTAGTTTCTA	GTCTCAGCAT	ACCAAAGCGT	CATCTGCAAT	TGAGAGCATT	GGATTGATGA	1740
TCTTCAAGGT	CCTTCCTGCT	CTAGCATTCA	CTGAATCTGC	TATTTTTGAC	ATATTGAATA	1800
ATCAGAAGCA	GCCAGTTTTA	GAATCTTATT	ATAGCAAAAG	TGGTAAAAAT	AATGAGCATA	1860
TACTATCAAT	GTGCATCTAT	GTCTTCTTAT	GTTTGAGTGA	GGATCCTGAT	ACATAAACCT	1920
TGGCTGATAA	TTTCTACTGA	AAAAAATCGT	AAGTATTAAA	GACACTCTTC	TGAAGATGTT	1980
CTCTCCAGAC	TCTGCTACAG	GCAATCATGA	GCAAGAGGGT	TGGCATCATC	GGAGCTGGAG	2040

TCAGTGGCTT	GGCTGCCATA	TGGTGCTGTC	TGGAGGAGGG	GCTGGAGCCC	ACTTGCTTTG	2100
AAAGGAGCGA	TGATGTTGGA	GGCCTGTGGA	AATTCTCAGT	GAGTGGCACA	TCATTAGAAC	2160
ACCAGTGGAA	GGAGATGGAT	TCCAATGCAA	ATCAAATCTG	ATCAGTTCTA	ATTCAGATTT	2220
AGAAGGCAGA	TCACAAAAGC	TCCAAATCTG	GAAAGTAAAA	TCTTACCTCT	CCAATCATAC	2280
TAATGCCCAA	AAAAACTATT	TCATACCAGC	AAAATTTGTC	CTGAAAAGGA	CATTTTCAGC	2340
TCATTAAACA	TCATCACCTG	CATGGTGAAA	TCCAGATCTC	CAAGCTGTAA	AGGGCACTAA	2400
TGTTGGTAAT	TAGTCAAAAA	TATACCATGG	GCTTCCCAGG	TAAGTGAAAC	AATTCTATTC	2460
TTTATTGCTC	TTAAATGCCA	GGAACACGAC	TAGAAAAGAG	ACAAACAAAC	CTGGACTGAG	2520
ATCCTGAGGT	CAGAAGTCCT	GAGTTCTAAT	TTCAACTTGT	AGGTTTTCTA	GGCAGATAAG	2580
ATTTCAGTCC	AGTTGCTTTT	GTTTCCCTGG	ACCTCAAATG	CTCATTTGTC	AAATGCAGAG	2640
GATATGATTC	TATAATTAAC	TTATGTCTAT	TGGGCAGATA	GAAATTATTA	TAGATGATGA	2700
TTGTGTGTGC	GGCTGTTGAA	TAGCCTATCA	GCTCCAAATC	CAGAGGGAAA	AATTATGGTC	2760
TTTGCCATTT	GGGCTCATTTG	TAGAAATAAT	ATAATTAGGA	AATAGTCCTT	GTAAACACAT	2820
TTTTTTTTTAA	ATTTCAAAGC	CAAGTTTGGA	GAAACTTCTA	GTTCTTCTGT	CCTGGATTTT	2880
CCAGCCATTG	TAATCAGTTG	TCGATGATAC	ATATTTGGCT	TGAAAACATA	TTCACATCAT	2940
TCATATTGTA	ACTACTTCCT	GTCCTGGTCT	CAGTTACTGC	TCTGCCTGCG	CCAATAGCCT	3000
CCTCCAATAG	AGTATATCAG	TGCTAACTTA	GAACACATTT	TTATTCTTCT	CCAAGCTTTT	3060
TTTAAAAAAA	ATTGTGGTTT	TGTAACCCTG	AAAGCACTCC	ATGAGATATA	AGGTCATTAA	3120
TTTTTATTTT	CCAGTAGGGG	GTAATCAAGA	GTAAATATTT	TTCAAGAATT	TAATTTTCCC	3180
TATTTACATT	TGCTCAGGGA	AATGTGGACA	GCTTAGAGTA	AATCATAAAA	TGGCTTTCTA	3240
CCATCTCCCT	AGTAACAATT	AAATGATGCT	TGAGCATCTA	TTCTGGTAGT	TTGTGCTAAG	3300
TACTGGGATG	ACAAATATGG	AATATAATCA	CTCCTTGTA	ATGGTTCCAT	TTCATTTGAT	3360
TAAGCAAGCC	ATAATATAAT	TCCGTAATCC	TTTGATAGCA	AATGGGCAAA	AACTCATTTG	3420
ATAGCCGAAC	CTCTTCTGAA	ATCGTAAGGT	TAAATACCGT	GAATTGGATC	AACATGAAGC	3480
TAAGTCTCAC	CTTCTGTTGC	ACGGCAGAAA	TTTTATTGCA	TTTGACAGAT	TGCTGCCCCA	3540
GATCTCACTA	GGAGTATTAT	GGAGCAAAAT	CCAAAAATGT	ACACATTCCA	AAATATATCT	3600
GGCCCTAAGA	CTTTTAAAAT	AAGAGATTAT	ATAACTACAA	CAACAAGATA	GACCTTGTC	3660

CCATCAATTC	AATGGACGAG	TGCCTCGAGC	GTTTAGAAGA	GGGTGGACTA	CGGAAATCTT	3720
AGTAGGTCAA	AGAAAACCTC	CCACAGGGAT	GACACTTAGC	CTTGAAGGAT	AACCCCAGAC	3780
AAGCAAAATA	GAGGACCACC	TGTGACACAA	CTCCTAGAGA	GTGCATTTCC	CAATAAAGTC	3840
TGCGAATGGC	ACTCCATAGG	CCTATGCAGT	CATTGGCAGT	GTGCCAGCAC	CAGGTTAAGA	3900
GAGACCAACA	ATCCATGAAA	GGCACAGAAA	AGGCAATGAA	CATGGTGTGT	GCAGAGAGGG	3960
ATCCATGAGT	TATCCAATAT	AGCCAGATCA	GAAAGTTTAC	TTAAGGAAGC	AATAATATGA	4020
TACAAAGATC	AGTAAGATTC	AAAGTTGGAT	TCTGAGTTAT	CCACAAGAGG	AAATTCTTCT	4080
TTTCCATAAG	GTCATGTCTA	TAAGCAAAAT	TCTACTCAA	GTCCTGGTGA	GGATATGGAC	4140
CCATACAAAT	ACTCAAAACT	TTAGCCTCCT	CCACATACCC	CAGCCCTTCC	TTCTTTTCTT	4200
AGAAAAGTTG	CTTGGCACAA	TATATAATCA	GAGAGGGATT	TTTTTTATGT	GTTACATAAG	4260
ACTTTATCTT	GTAAGCCTTT	TTTAGAAGGT	GTTCTAGCAG	ACAGAAACGT	GGTAATTCTG	4320
AACTTTTCAC	TATTTGCTTT	TTCTGAGAAA	TGAAAACCAA	ATGGGATTTA	AATACTAGCA	4380
GGCTGAATGT	GTGTTTTAAG	TTTCATCCAC	TCCTAAATAG	GGCCTCGTGT	CCTCAAAAGA	4440
TTTCATTACT	GCTGTAATAA	GAAGTTGCTC	AACAGCCAGG	TGCGGTGGCT	CATGCCTATA	4500
ATCCCAGCAC	TTTGGGAGGC	CAAAGCGGGT	GGATCACGAG	AGGTCAGGAG	TTCAAGATCA	4560
GCCTGGCCAA	CACAGAAAAA	CCCCATCTCT	ACTAAAAACA	CAAAAATTAG	CCAGGTGTGG	4620
TGGTGGGTGC	CTGTAATCCC	AGCTACTCAG	GAGGCTGAGG	CAGGAGAATC	TCTTGAACCC	4680
AGGAGGCAGA	GGTTGCAGTG	ACCTCAGATC	ATGCCACTGC	ACTCCAGCCT	GGGCAATAGA	4740
GAAAGACTCC	ATTAAAAAAA	AAAAAATGC	TTACCAATAG	GTTAGTAGCA	TTTTGATTGC	4800
AAAAGCTGAA	GCCAGGACTA	TTTGAACTTT	TTCCCACTCA	TTTATTCCTT	TGTTCAATTCA	4860
ATGAATACAT	ACTGTGTACT	TTATGTGTAG	GGTACTATAT	TAAGCATAAG	CTGCAGATAA	4920
GAGGCCAGCC	AGCACTTTAA	AAGCCGTGAG	AAAACAAGTA	TCAGAATAAC	TATAAGTGAC	4980
TATATAATTA	GGGCAATAAG	GATAATGGGA	CCTTAGTAAA	ACTAAAGATG	ATTTGGCAGT	5040
AGCTGAGAGG	GAAGGTAAAG	AAAGCCATGA	CAAAGTTGAA	GGCAACTTTT	GAGCATATTT	5100
CAAGGGCATA	TTTAGACAAG	GAGATATGGG	ACTCATAAGC	AGAGCTGGAA	TAGGAAAGAA	5160
GATCAAGGTA	AACTGCTTAG	ATGCATGTAC	AACATTCTGA	AATTAACCTC	TGACTTTGCC	5220
CTCAAGTTAC	TTATGTTCTC	GTGGGAAAGA	TGAGAGATGA	ACACGGTTAT	CATCCAAGAC	5280

AGATGGTGCC	CACAGCTGCT	TAGATCTCTG	GTTCCAGGGT	AAAGCTCCCT	CAGCTAGAGG	5340
CAGAGTCAAA	GTTGAATTTT	CTCCTTACTG	GCTCAAACCA	CACCTCATAT	TGAAATAATA	5400
AAAATGCATG	CTCCCTGGAG	CAACTGACTT	GTTATCTAAT	ACATTTGCTT	TTTTGTGTTC	5460
ACTTGAGAG	CAGTCTTTTC	GGAAAAATTC	CAAGGAGCTG	TAGTGTACAT	ACTCTTCTCT	5520
CCTGGTGTTA	TAATTGGCTG	AGGTCAAGGG	GCAAAAAAGC	AGAGATTCAT	TCAAGATGGA	5580
AATATTCCAA	GGCCTTAGCA	TCTGTTTCCC	AGAACAGAGT	CTTACATTCT	TTAACCAGGC	5640
TCCATCCCAC	AGTTCAGCCC	TGCCTCCTTT	CAACAGGCAG	CTGAAAAAAC	CTCCTTCCCA	5700
CCTCTCCTTC	TCACAACCAT	CAGTAGAAGG	CGCTAGCTGT	GGGTGAAAGG	GAAGCACTCA	5760
GCCTGCCAAA	CTGCTGGACA	TGAGCCTTCA	CCCTTTTTTCT	GACCTCCACA	AAAATTTTAA	5820
AAAGTTTAAA	TTCCTGTGCT	TCCACGCTTA	TGAGAAATAC	AGCAACCATG	AATAGAGGAA	5880
GATTATGTTT	TCAACTTGAG	AAAAAATACT	GAGGCTTTGG	GCAGCCCCCC	ACTTCCCCAC	5940
GGGGACACAA	TCCTCTCAAC	CCTTTCCAGC	ACTTTTTGTT	TCCCTCTTCC	AGAGGTCATC	6000
TGGTGTGAGA	GGGAGATACA	CATCTTGAAT	CCAGCAGCAA	CGTGACATTC	CATCTCTTTC	6060
CCCCCATTGC	ACAAGAGTCC	CTTCCGGACC	TCGGGAAGCA	GAAGCTGCCA	GCTCTGAAAT	6120
GTATTTTCAA	GGCAGCACAT	TGTGTGCACT	TTTACCCTAC	CCTCACAACT	GAGAGGAAAT	6180
GTTTATTTTC	AATTTAGCTT	TTGACTGCTT	CTAAAAATA	AGCCACTTTT	CAATTACACA	6240
GAGGCTTTAA	AATGAAGTGC	CAAGATTTAA	CACATGTTCT	AAGGGCTCTG	GTTTCCTGTG	6300
TTTCTTTGGT	GAGGAGTGAA	GTCCAGCAAC	TGGTGAGCCA	AAGAATAGGA	TTCATTTACA	6360
ACAGAGCAGT	GGTTCTCAAA	GTGTGGTTCC	TAAACCAGCC	ACATCAGCAT	CACCAGGAAC	6420
TTGATAGAAA	TGCAAACCAC	CCCAGACTCC	ACCCAGACA	GATTGAATCC	GAAATTCTAA	6480
GAATAGGGCC	CAAGAATCTA	CGGTCTAGGG	AGCTTCCAGG	CGATTCTCAT	TACGCCAAAG	6540
CTGGGAAACC	ACTGCAATAT	TGGGTTGTTG	CCAGTGAAGA	GTTTGCTAAA	CTCCAAAAGC	6600
AAATAAATAG	GCTAGAAGTC	AGAGCCTCTT	CTAGACAGTT	TTGTTTTTTT	TTTTTTTTTT	6660
AACCTGAGTA	TAAGATCAGA	ACCAGTGGTG	GCACAGGAGA	AAGCAAAAAC	CACTAAGTGG	6720
CTATAAAGAC	AGAGCTAACA	CTGAGGGTAA	TTACAGTAAG	AGGATTACAC	TGGAAAGAGC	6780
TCCAGTTCTG	TGCCAGGTTA	CGCGAAGGGC	TTTCCATTCC	TTATCTTACT	GAGAGCTTTT	6840
AATTTTTGTT	TACGCTTTTA	AACATGAAAA	GGGTTTTAGT	CAACCAAGAA	TTGAACCACT	6900

GTGTTCACTG AAGGGAACAC AATTCTTGGC TTTCTCTTTA AGCTTTCTTA TTCTCCCTAG	6960
GACCACACAG AAGAAGGCAG AGCCAGCATT TACCAGTCTG TATTCACAAA CTCTTCCAAA	7020
GAAATGATGT GCTTTCCAGA CTTCCCTTAT CCGGATGATT ACCCAAATA TATACACCAC	7080
AGCAAGCTCC AGGAATATAT AAAGACATAT GCTCAAAAGA AGGATCTTTT AAGATACATA	7140
TAGTTTGAGG TAGGGGTCTC ATAACCTGTA CTGTTGAAAT TAAGATATGT GTGGGTTAGA	7200
GAAAAAGGAG GCAGCAAACT ATTATAAAAA TTAGAGCCAA ATGTTTGGGC ACCTCAGTAA	7260
TCAAATGTTG GGTCTGATTA TAAAGCATTG ATGCATTGAT TTTTCTCTC CTAGACTTAC	7320
TAGTTCACTA GTCTCTGAGA GCTTTCAGAC TACCTTAGAA AATGGAGGCA GCTAGCCCAT	7380
CATTGTCCAC TTTCCACCCT CATGCTCTGA TGTTTTGGAA ATAATCCAAA ATGCTTTAGT	7440
ATATATTAGG AATTTTGTCA GTTCAATGCC AATGAGTTGT GGTTCAAAA ACCAGAGCAT	7500
TTGGTAGGGT TTCTCCCATT ACATTATGAA AAGGTTAACA ACTTAAATGG GAAATATAGT	7560
CATTGCCCCC ATCTTTACCC ACTCAGTTCA TTAGTTTTTT TATTAAAAAG GTGAGATTTC	7620
AGCATTGTTT CTGCGAGAAT AATGTTTTAC ATTTATTTGG GACTCTTTAT TGAGCATTTT	7680
TGTCTGTATG TTTGGAAGTC TTAACCTCAA TTAAGTCTG CTAAATGCAG AACACTTGCA	7740
TATAGTGGGA AAAACAATCA GCAAAATTAT GAACCATGGT GATATTTACA TCATTATTTT	7800
ACCTGGAGTA GCCCCAAATG TATAGTTAAA ATAAAATTTT CCAATAGTCA TTTTATTCCA	7860
TTCATTCAAT ACATTCAATTT GCTTCCATTA TGGTGTTAAT ATCAACAAAC ATTAATGAAG	7920
TTCTTATTGT GTGCTTGCAT TGTGCTATGT GTTATATGTA AAAGAAAAAG AGGTCTAAGA	7980
CTTAGCTCTC AAGAAGTTAT TTCAAATAA ATATGTAAAG AGTAAGTAA AAGATTCCAG	8040
TAACAATTTT AATCAAAGAG AAAATTTTTT AAAGCTCTTT ATGATTTGTT TATAAATAAA	8100
ACAATGCTAT GGAGATCATG AAGCAAGAGG CAACACTTTG GGGGAAGGTA TTTTCTAGAG	8160
GAGGTAAAAT TTAGTTGTAT TTAGTAGGTG TTTTAGATAA ATGAGTGGCA TGAGTAAAAT	8220
TAGAGAGGTG GGAAATGCC CTGCTCATTT GGAGAACAGT GGGCAAACCA AGTTGGTTAG	8280
GAGGGAGATA TATATGCTAG GATGAGATAT GGCCACATAT ATCAGTAAAC TAGTGTGTAC	8340
TGTGACTTTG AAAAATAGAG GATTATTTTG CAACCATGTA AAAGAAGTCC AAAGAAGGGA	8400
CATCCAGAGC TTATGTGATG GCACCAAAGT TATCAAAGAT TCAGCTTCAC CCATCTTAGC	8460
ACGTGGCCTA CATCATGACG TTTGCCTTGT GGTGCAAAAC AGTTGCTGAA GCTTGAGCCG	8520

TCACATCTGC	CTTCTAGCAA	AAAAAAAAAA	AAAGTAAAGA	ATGAAGGGCA	AAGGGATGTT	8580
CTCTCAGCTG	AATCAGCTCC	CCTTTTACAA	ATTCTCCTGA	AAAACTGTC	CAACATTGCT	8640
TATATCTCAC	AGGCCACCCT	AGTTGCACAG	GAACCTGGAA	AATGCATCCC	TTTTCTGTGT	8700
ATGTTGTCGC	TCCAAACAAA	ATCAGGGTTC	TGTTAGTAAG	AATGAAGGGA	GAATGGACAT	8760
TAGGGAAGCA	ATTTGCAGAA	TATGTTCCAG	AAAAGTCTGT	GGGAATAACA	GAAAATAAAA	8820
CTAAAAGAGT	AAATTGGAAC	AAAATTGTAT	GGACTTAATA	GTAATCGCAT	TCAAATGTA	8880
GAATAAGTTT	TAGAGGCTGT	GAAGTAACAG	AAATTGAGCA	GTGAATTGAG	CAGAGAAATT	8940
GAGAAATGAA	TATAGTCCTT	CAGGAAGATT	AATCTGACAA	GCAGGACAAA	GGATGGCTTG	9000
TAGGAAATGG	GAGGCTGAAG	ACAGGCTAGG	TATAGGTTCT	TGCCGTAGTC	CATGCAAGGG	9060
AGTGATAAGG	ACTTGAATGA	AGGCAGTGTT	AGCAATCATG	GAAAGAAAGC	GTGAGATTGG	9120
GAGATAAATA	CTGTTTAAAC	ATGAGGCAAG	GATGGAGAAA	TAACAAGGAA	AACAAGTCAT	9180
GGATTTGAAG	CATAAGTGGC	TGGGAGTTTC	ATGTCATCAT	TCAAAGAAAT	AAGAAAGTCA	9240
GAAGCCAGTT	TCAAAGGAAA	TTTAAGTAGG	TCAATCAAAA	CCTGCTACAT	ATGAGGAAGT	9300
ATTAGGTGGC	CCTCCAGATG	GAAAGGTCAA	GCTAAACTGG	ATAGAAGAGA	GACCAAGGAT	9360
AGATGTATTT	GTATATTCAT	ACCACAAAAC	TTGCTAATTT	TTTTTTTTTT	TTTTTGAGAC	9420
GGAGTCTCGC	TCTGTCGCCC	AGGCTGGAGT	GCAGTGGCGC	AATCTCGGCT	CACTGCAACC	9480
TCCGCCTCCC	GGGTTACAC	CATTCTCCTG	CCTCAGCCTC	CTGAGTAGCT	GGGACTACAG	9540
GCGCCCGCCA	CCACGCCCGG	CTAATTTTTT	GTATTTTTAG	TAGAGACGGG	TTTCACCATG	9600
CTAGCCAGGA	TGGTCTTGAT	CTCCTGACCT	CGTCATCCAC	CCGCCTCGGC	CTCCCAAAGT	9660
GCTGGGATTA	CAGGCATAAG	CCACCGCGCC	TGGCCGTAAA	GTTGCTATAT	TTCTAAGATA	9720
AGAGTATTTA	TGCAGAGCAA	AAGAGATGCC	AACGATCAAA	CCTTGAGATA	TTCCCATACT	9780
TATTGAGTAG	ATGGAAGATG	AGGTCAGAAA	AGGAGGAAGC	CATGTCAGTA	GAGGGTAGCC	9840
ATAAGAAAAT	AACACAGATT	TGTTATATGA	CATCATTCAC	AAAAATATTC	AGTGTGATTT	9900
ACCCCTAAAT	CAACTAACTT	GATGTCAAAA	AGTAAATGTA	CTCCAGTGAG	TAATTTTTCT	9960
TGTGAGATTC	AAAGACTCAC	TGAAGATTCA	CTGTGACTCC	AATTTTACTA	TCTTTCTATA	10020
CATTTCTGAA	TGACCAAGAG	AGCTCGTAAC	AATTATTTCC	TCCACAGAAA	CAAGGCAAGA	10080
AGGAAAAAAA	CTTTCACATG	TAGAATTATA	AATGGAAAAA	TAAATTTTCT	AGTTTTCTTA	10140

AAGACCCTGG	TTTCCGGTAT	AAAGAAATGT	CCCAGCTTCT	TAGTCACGGG	CCAATGGGTT	10200
GTTGTTACTG	AAAAGGATGG	GAAACAGGAA	TCTACTATTT	TTGATGCTGT	AATGATTTGT	10260
TCAGGACATC	ACGTATACCC	CAATCTGCCA	ACGGATTTCCT	TTCCTGGTAA	GTTTGGAAAA	10320
TATATAATAA	TCTAGGGACT	TATATGCAAA	CATCAAGAGT	TAGAAACATA	TCTTTCTATA	10380
GGTATTACAT	AATGATTATT	CTTAGATTTT	AAAAGAAAAA	AATTAAGTTT	AATGATAGGA	10440
TATAGTAATA	AATAGCCTCA	TAAGTCCTTA	TGTTAAATA	ATCAAGGACT	GCAAGCCAGA	10500
GATCAGACAA	ACACAAGTTC	CTGTGTTACA	GACAGTAACT	CAAATATAAG	TTCTAACAGC	10560
ACACGGGGTC	TCCGAGCACA	GTTACATTAA	AAAAAAGTAG	AGTCCAAGT	CCAAATGGTT	10620
TAAAGAAAGA	CACGTTTACT	TATGTTATTT	ATAGGAGACT	CCTAGGTTTC	TAATTTTCATC	10680
TTCATCCACA	ATTTGCAAAT	AACTTTTAGA	AATCTCAGTG	ATTTGTGTGT	GGGTACACAC	10740
ATGGGTGTGT	GTATAGCAGC	ATACTTCATT	ACCATCCGAA	AGTGGCAAAC	CTCAAATAAA	10800
TACAATATAC	ATGGAGGCTT	CCTTCCATTT	TTCCTTCCTT	CCTTGCCACA	GGAACACAAT	10860
CTACTCAAAG	ATATTAGAGT	TTCCATGTCT	AGGTATGATG	TCCATAGGCC	GAGGAAAATT	10920
AAAGAGTGAA	GGTTCAGGAG	GAATATAAGA	TTAAACTCT	TAATGTTAAC	GGGCAGCATA	10980
TTTAATGTTT	ATGAGCATGG	GATCAGAACA	CCTGGCCTCA	ACTTACTATT	CCACTAGTTC	11040
CTTACCACTT	AACTTCTTTG	TCTCAATTTT	CTCTTCTTTT	AAAATAGGGA	CAATAGCCCA	11100
CCATGCAGGG	ATGTTATCAA	GATTAAATAG	TTAAACGTG	TAAAGCATTT	ATCAGAGGAT	11160
CTAGCCCACA	GAGTTAACTT	AATAAATATT	AACCATTATT	ATTATCGAAA	CATACATTCT	11220
CATGCCTTAA	GATTTTTTAA	GGAATAAAA	GTAAGTTTTA	GGGGGCTTAA	TGTCAAAAAA	11280
TGCTAAATGG	ATAAATGCAC	TTCAACTAGG	GAATTTTTTA	ATTACAAGT	ATAATAGGTT	11340
TAAAAAGACA	CAAAGAAAAC	ATCTTCATAA	TTTCTGAAAA	TCAGTTCAA	CACTTGCCA	11400
TGTTCCACTT	AGGCCTGGAC	CAGTTTCGAG	GCAACTACCT	CCATAGCCGG	GATTATAAGA	11460
ATCCAGAAGC	CTTCAAGGGG	AAGAGGGTCC	TCGTGATTGG	TCTGGGGAAT	TCGGGATCTG	11520
ACATTGCTGT	TGAGCTCAGC	CGTCTGGCTA	CACAGGTACA	TGACGTAAAG	GTTTTGGGAA	11580
ATAAACCTAA	GGTAGGGCTG	TGCTACTAAA	TCAGTAGCCA	AGGCACAGAG	GATGGTACTT	11640
CTATGTCACA	CCACAAGAGA	TCCACCTCTT	CTATGTGGCC	CTTCAAATCA	AGGAGGACTT	11700
GAGACATCCT	CCATGTGAAG	CCAGGTAATG	TGGCCCGTGC	TAGTAAGGAA	GTACATTCCA	11760

CTGAATCCAG	AAGTAAGTGC	ATGAGTGCGT	GTATGTACAG	ATGAGTGTGT	ATGTGTGTAT	11820
TTCTTGTTTT	CATTTTATAT	TCTGATCACC	TCCAAATAGA	CTAGTTCCTG	GTCAGGCTTA	11880
ATCTTTATTT	ATTTAACAGT	ATTTATTATA	ACGTATCATG	CAAAAAGCAC	TGTGTTTACC	11940
ACTCTGAAGT	TCTGAAAGAT	ATGCATGACT	TGGTATTTAC	TAACATTAAT	TCAATCAACA	12000
GCAGATGCTC	AACAAATATT	GGGCACTTAC	TATGCTTACT	ATGTGTCAGA	ACTATGATAA	12060
ACTAAAATA	AATGCATAAA	TAAGTTAGAC	TAGTTCCTGA	CTTCAAGAAA	GAGTCAATGG	12120
ATGGAGATGG	AGTTGACAGG	TACACACAGA	CTATCACCAG	AGGAGATGGT	GAGTCTTCCA	12180
GTAGAATTAG	GTGTGGCAAT	AGCAACACAG	GGAAAAGAGA	ATCTAACTTA	GCCTGGATGA	12240
GGTCAAGGAA	GACTTCCCAG	AGGACTCCAA	GCTAAATCAT	GTATCATCGA	TAGACCCTAA	12300
AGAAACAACA	TATTTTTAAG	AAAACAGGTT	CTCAATAAAT	AAATTCTTAA	ATGGATGTAA	12360
ATAAACCTT	AATTTTTTAA	ACTAAAAATT	CCCTTCAGTT	ATCACAAAGT	TAAAGTCTAT	12420
TTTGCAAAGA	CGGTAAAATA	GATAAGCAGC	CAGACTCATC	TCAGGGCTGA	GGCGGTTGCC	12480
ATGGTTTGGG	TTGCTCAGGA	GAAGTCCTTG	GGGTATGTGT	ATAGGGAGAA	CTGGAAAAGG	12540
CAACCAGAGA	CAGAGAACAG	AATTAAATCC	TTGACATCTC	GTCAGCCTAA	TTTCAGCTAG	12600
AGATTTAGCT	ACACTTTTCC	CACACCTAGT	CCACTATCAC	CAGCCACAAC	CACTGGGGCT	12660
CACTGGATCA	TCTGGTCCCT	ACCAGACTTG	CCATCTTAGT	CTATGAGTAT	GTGAAGATTA	12720
AACCATCACA	GTTGAACACA	GAGCCCTGTT	GTTCTTAGAG	TGATGATTCT	AATCCTTTCA	12780
ACAACCTACAC	ACCAGCCCTC	AGGGGCAGTG	AAAGAATCCT	GTCTCTACTA	GTTTAAATTT	12840
TAGACTTTAA	AAAAAATTTT	TTTTATTTTA	AGTTCTGGGA	TACATGTACA	GAACATGCAT	12900
AGGTCTGCAC	ATGCCATGGT	GGTTTGCTGC	ACCTATCAAC	CCTTCATCTA	GATTTTAAGC	12960
CCCACATGCA	TTAGGTATTT	GTCTTAATGC	TCTCCCTCCC	CTAGCCCTCC	ATCCCCCGA	13020
CAGGCCTTGG	TGTGTGTTGT	TCCCCCTCCT	GTGTCCATGT	GTTCTCATGA	TTCAACTCCT	13080
GCTTATGAGT	GAGAACATGC	AGTGTTTCGGT	TTTCTGTTCC	TGTGTTAGTT	TGCTGAGGAT	13140
GATGGTTTCC	AGCTTCATCC	ATGTCCCTGC	AAAGGACATG	AACTCATTCT	TTTTTATGGC	13200
TGCTAGACAA	CTTATTTAGA	CTCGCCTTTT	AAAAGTGTTT	CTACTTGGAT	ATTGAGGAAA	13260
ATGCACGGAA	GTGCCCCAAG	AAGTGTGTTG	TGTTTGCTTA	TTTCTTACAG	AGTAATGCTG	13320
AAATCTGTGT	TGCTTTTCCC	CACCAGGTCA	TTATCAGTAC	CAGAAGTGCT	TCCTGGGTCA	13380

TGAGTCGGGT	CTGGGATGAT	GGCTATCCTT	GGGATATGAT	GTATGTTACC	CGCTTTGCAT	13440
CCTTTCTCCG	GAATGTCCTT	CCTTCATTCA	TCTCTGACTG	GTTATATGTC	CAGAAGATGA	13500
ACACGTGGTT	TAAGCATGAG	AACTATGGCC	TGATGCCTTT	AAATGGGTAC	TTAAAAATGG	13560
AAATTTTTTT	TATTCAAAAA	AGGGGGGCAC	TCATTTAATG	AATTTATTCT	CTCTAGAACT	13620
TACTTTTGTT	GTCTCATTGA	GCCTAGAAAC	ATTAAACTCA	AGGTTTCACA	GGTGACGGAA	13680
TATGCCCAGA	GACCACGTAT	GGCTTGAAAA	ACTTATTGAA	ATTAGTCCAG	TACAGAAAAGG	13740
GTATGGAAAA	ATCTGAAATG	GAGATGACGC	AGGCAGATAA	ATCACCTGA	CATGCATGAT	13800
GCATTTGTGG	TGGCTACAAG	CTATAGCATA	GAACTTTGAG	GACTGAACAA	ACTCAAATTG	13860
GTTTTTGGAA	GAATATCTTG	TCCGTGCTTA	TGGGTGTATG	AAGACATCAA	TAATAATACT	13920
TGCTTCTCAA	GATGGTTGTG	GTATTCAATA	ATATAAAAAAT	ATAAAAAATTG	CTTTCTAAAT	13980
GATAAAGCTT	TAAAAAAATT	GGTTCTTCTT	AGTCTCAATT	TTTCTAATGT	GCTTCAAAGG	14040
AGCAAATAAC	AAAATAGTGT	TAATCAACAT	GTCTCAGCAA	GTAGGAAGTC	TCAAACAAAA	14100
AGTGCACACT	TCCTCCACCC	CTGAAATGTT	GACATTTTTG	CAGAACCATC	AGGAGGCATG	14160
GAACACATAA	AGTAATGGAG	AGTCACAACT	AACGTGGCCT	GTAAGATTAG	TCAGATTTCAT	14220
TTATTTACTT	CTTTATAGAG	ACAGGGCCCA	ACATTTACTA	ATTAGGAAGT	CATTCCAGGT	14280
AGAAGAATCA	GCATATCAAT	AGAAAAAAG	AATATTTAAG	TTGGTAAGAA	AAGAAAGAAT	14340
TGAGAAATTT	TATCTCCTGG	CCCATGCTAG	CCAAAAAGTT	TCATTGTGTT	TAGAGAAAGA	14400
TGGTAAGAAA	AAGGAGGAAC	TGTAAATCAA	AAGAGCAAAT	GCCAGATTTA	GGAGCTAAAC	14460
TGTCAGTCCA	AAGCACTTAT	ACTACCAAGT	CTTGCAGGCT	GCTATAACCC	TTAAAAATAT	14520
GTTGATTTTA	TGCATTTAAA	ATTATGTTTA	ACACTGTGGC	TTGCTTGACA	GTAGAGGGTG	14580
GGAGGAGGAA	GAGAATCAGA	AAAAAATACT	TATCAGGTAC	TATGCTTATT	ACCCAGGCGA	14640
CAAAATTATC	TATACACCAA	ACCCCTGTGA	CACACAATTT	ACTTATATAA	CAAACCATGG	14700
ACCCCCAAAC	CTAAAATAAA	AGTTTTTAAA	AATTATGTTT	AATATAGTAA	GTCCCATAGC	14760
TTGAGCTGGT	TAAGATTTTT	TATCTTGTA	GAGTAACTAT	AAATTATATT	TTGGCCTTGC	14820
CATTTAGACA	ATTAAAACAT	AGTTTTAGAA	ATTCATTCAT	TCTGAAAAC	AAGCTTCCTT	14880
TTGGAAAGGG	TTCCAATTAC	CCTAAGTTTC	TGGAGGGAGA	AAGGGGGAGG	AAAAACAGGT	14940
TTCATTGTGG	TCTATGTTTT	GCTACCTTGT	AAGGTAAAAAG	AAGAGGTTGC	AGGATTAGAT	15000

AAACAGAAAA	TGATGTGGAA	GTATAGAGAC	AAATTTTCAGG	ATTTACAAGG	TTTCTTTGTG	15060
TCTGAGATAC	TTGCAGGAAA	TTCCGGAATC	TCAAAGGAAA	CTTAAATCAA	AATGAAATAT	15120
ATTGTCTCTGA	AAAATATTAT	TCCTAGAATT	TTGGCAACTA	AAATGCAATA	TCAAAGTTGT	15180
TACACTTTTTT	TGTGGACACA	GCTGATGAAA	GAAAACCAAA	CATGGCAATA	AAACTTCCCA	15240
CCACTGCAAG	TCTGATTTCT	CCATGTAAGA	CAAGACGTTA	AAGTTATGAT	AATAGTGCAC	15300
TTATAACAAC	AGTGCTTGCA	TGTGCCAGGA	ACTGTTTTAA	GTGCTTTAAG	GATAATTGAT	15360
CATTTAATTT	TCACAACAAC	CTATGAGGAA	GATTCCATCA	TCATCCCCAT	TTTACACATA	15420
AAGAAACAAA	TACAGAAAAG	TAACAAC TAG	TAAGAGATGG	AGCTAGGTTA	TGAACCTGGG	15480
CCATCTGCTT	CCAGAGTTGG	CGTTCTTAAC	CACTTTAGTA	TGTCTATAAA	TTAGTTTTAG	15540
TCTCATTTAG	GAAAGGAATT	GCCATGAGAG	AAGAGAGTCA	GTGGCACTCA	TGCTGATGTT	15600
TAAGTGCTTG	ATGTTATTTT	AATGTTATGG	GCTGTTGCAG	GTATTTCTTG	GAAATGAGCT	15660
ATTTACAGCA	AGGGTGTTTG	CCTCTCATTG	CTGTAGTTCC	CTGAGAAAAG	AGCCTGTGTT	15720
CAATGATGAG	CTCCCATCCC	GCATCCTGTG	TGGCACTCTG	TCCATCAAGC	CCAGTGTGAA	15780
GGAGTTCACG	GAAACCTCAG	CTGTGTTTGA	GGATGGGACC	ATGTTTGAGG	CTATCGACTC	15840
TGTCATCTTT	GCAACAGGCT	ATGATTATTC	CTACCCCTTC	CTTGATGAGA	CCATCATGAA	15900
AAGCAGAAAC	AATGAGGTTA	CCTTGTTTAA	AGGCATCTTC	CCCCCACTAA	TGGAGAAGCC	15960
AACCTTGGCT	GTGATTGGCT	TGGTTCAGTC	CCTTGGAGCT	GCCATCCCCA	CAGCAGACCT	16020
GCAAGCCTGG	TGGGCTGCTA	AAGTATTTGC	AAGTAGGTGG	GCCATTCTGT	CTTTCATTCA	16080
TTTTATCAAT	GAACATTTAC	TGAACACCTG	CTATATGCAA	AGCACTGTGC	TAGGGATACA	16140
ATGAGAACAA	GACAAACATG	TTCTTGACC	TCTCAAGGCT	TAAAATGGGG	TGTGGGGGAT	16200
GCCATAATAG	GGGAAATTTG	GGGGGGTTCT	AGTGAGGGGA	GTTGGACTGT	TGCACAGAGC	16260
AAACAGTATA	CAGGAAGTCA	TAAAGGTGAG	GGAAAGCATG	AAATGTGTAA	GGACCCAGAA	16320
ACATTTTGGT	GGAAGGGAAT	ATAAAGCAGA	GGCAGGGAGT	GGCAAGAAAT	ATAGGTTTAT	16380
AAGCCACGTT	AAAGAGCTTA	AACTTCTCAT	AGGGATTAAG	GACTTCGCAA	GATTTTAAGC	16440
AAGAAAAAAA	TAGCAGAGGA	TAACTGCAAT	GTCAGGCTAC	ATTATAAAGA	TTGAAGGGC	16500
CCTGGTGAGG	GTTGGAGGTG	TGCCAGAAAC	CTCACTGGTG	TCAACTTCTG	TCAGAATAAC	16560
AAAGTCAGGC	CACTCTGATT	CTCATGACAA	TCTTCTTCTT	CTCTCCCTCT	ACTCTAGACC	16620

TCATGGTCTC	CAGGGGCTAC	AAGTATGCTT	ATGTGAGGAA	ATCAAGAATA	TGAGGATTAC	16680
ATGGAGAAAAG	GCAATGTCTC	AAATATATTA	ATTTACTCCA	GTCATACTGA	ATATTATCAT	16740
TATTATTGAA	AAGTGTCTT	TTATTCAGGT	ATTCTCCAAA	ATATTGACCA	ATATAGGTAT	16800
AACTTACCTA	ACATAACTAA	TCCATAAAAA	CTTACACTAT	TGGTAATTAA	CAAACCATTA	16860
CAATCATGGA	ATATATGTAT	ATATATTGTC	TAAAACTTTG	TAGATAAATA	AATTTCTATT	16920
TCAAATACAC	CATGAAAGAT	CATCATTTAA	ATAAACCCCA	TCATGAAATC	TTTTGTAAAG	16980
GTGCTCCCTG	CAAATACTT	CTATTGCCTT	TTTCCTTCGA	AAGGCACAAC	AATGCCAAGA	17040
GCCTGGGGTA	TTATGAGAAG	ACTGGATATA	GTCATAAAC	CTAAGAAATT	TACATGAAGC	17100
AAATGGTATC	ATTTATTTAT	TCAGCAAATA	CTTACTGAAC	ACCTACTATG	TGTCAGGCTC	17160
TAACCTGGCA	CTTAGGACAC	AACAACAAAC	GAAGCAGAAC	AAAATTCTGG	CCTCTTACTT	17220
TCTAGCAGGG	TGTCCAGCCA	ATATCAATCA	TAGGGTACTA	CCAGGTTGAC	ATAAGACACT	17280
AACGATGACT	GGGAAATATT	CATGCACTGC	AAATTTTAGA	GTAACTTTCT	TCCACTGTTA	17340
CAAAGGCAAA	TAAGCTACCA	TCACCAGTTA	AAAGAAGTTG	CATTGATGTA	GTGAAATTCA	17400
CAAAAAGCTA	AAACTTGTCT	GCTGCCCTT	AAAACACCTT	GCATAGTTGC	AGAAGATGTT	17460
TAAAATCCTA	TGCTTCCTTC	CATTACCTCA	TTTAAAATGG	CAGAAACCTT	AAAGGGAAC	17520
GTTTTACCAG	ATTCTTTCTT	CAGAGAAGTT	TTAGGAAAAG	GATACAGAAA	AAAAAGGAAG	17580
AAATTATTAA	GCTATTATAT	GCATGAAGTG	TACTGAGCAC	ATATGTTGAG	GATTAGGTCC	17640
TCTATAATGT	TACCGAAATA	AGAGACTGAG	TGATTTGAAG	CTACAAATGT	CTCTGCTGTC	17700
ACTATCTCAC	TACAGGCCAG	CTTTTCCAAT	TCCCAAAGGT	TCATTAACTT	TTCAGATCTT	17760
TGTTTCTATG	AACTGGTATT	TTGCTAAAGA	TATCAAAGAC	ATCTCCAGCT	CCTCTTAATA	17820
CAAAGTTTTT	CAGGAATACA	GTTTATAAAA	ACCAAATGAT	TTCCATCATA	TGTCATTATA	17880
TATTTCTGAT	TTGTGTTTTT	CAATATTTTT	CTCTTCATTT	CTTTTCTAGA	CTCATGTACC	17940
CTGCCAACCA	CGAATGAAAT	GATGGATGAC	ACTGATGAGA	AAATGGGGAA	AAAACCTCAAG	18000
TGGTAAGCAG	CTAACTGTAC	TTGCTAATAG	AGCAAGTTCC	TAAAATGTGC	CTTTATGTGT	18060
AGAAAAACAT	TAATATGCTT	TAATATTGTC	ATTAGTCAGA	GTTTACATTT	TCTGAACACT	18120
TGCAATAATC	AAAAAATGTT	TAGATAGTAA	ACAGTCATCA	CACTTCTCTT	GTGTAACCTCA	18180
AGAATAGAGG	TTTTCTATCA	GGGATAATTT	TGCCCTCCAG	GTGACATATG	GCAAAATCTG	18240

GAGACACTTT	TGGTCATTGT	GAGTGGAGAG	GGCATGCTAT	CAGCATCTTA	TGAGTAGAGA	18300
ACAGGGATTC	TGCTAACCAT	CCAACAATGC	AGAGCACAGT	TCACCAAAAC	AATTATCTGG	18360
CTCAAAATGT	CAATAGTGCT	GAGGTTAAGA	AACAACTCTA	TAAATGACTA	CAGTTGACCT	18420
TTGAACAACA	CAGGTTTGAA	TTATATGGGT	CCACTTATAC	ATGGATTTTT	TCAATTAACA	18480
TAATGCAGAT	TGGGCATGGT	GGCTCACGCC	TGTAATCCCA	GCACTTTGTG	AGGCTGAGGC	18540
GGGCGGATTA	CCTGAGGTCA	GGAGTTCGAG	ACCAGTCTGG	CCAACATGGT	AAAACCCTGT	18600
CTCTACTAAA	AATACAAAAA	AAATTAGTCG	AGTGTGGTGG	TGTGCACCTG	TAATCCCAGC	18660
TACTCGGGAG	GCTGAGGCAG	GGGAATTGCT	TGAATCAGGG	AGGTGGAGGT	TGCAGTGAGC	18720
CAAGATCGCG	CCACTGCACT	CTAGCCTAGG	TGACAGAGTG	AGACTCCATC	TCACAAAAAA	18780
AAAAAAAAAA	AATGCAATTT	TTTGGAGATT	TGCAGCAATT	TAAAACTCA	AGGCCAGGCG	18840
CGGTGGCTCA	CGCCTGTAAT	CCCAGCACTT	CGAGAGGCCG	AGGCGGGTGG	ATCATGAGTT	18900
CAGGAGATCA	AGACCATCCT	GGCTAACATG	GTGAAACCCC	GTCTCTACTA	AAAATACACC	18960
AAAATTAGCC	GGGCGTGGTG	GCGGGTGCCT	GTAGTCCCAG	CTACTCGGGA	CGCTGAGGCA	19020
GGAGAATGGC	GTGAACCCAG	GAGGCGGACT	TGCAGTGAGC	CCAGATTGTG	CCACTGCACT	19080
CCAGTCTGGG	CAACAGAGTG	AGACTCCGTC	TCAAAAAAAA	AAAACCTCAA	AGATGAATTG	19140
TGTAGCCTAG	AAATATTTTA	AAAAATTAAG	AAAAAGATGC	CATGTATAAA	ATATTTGTAG	19200
ATACTAGTCT	ATTTTATCAT	TTACTACCAT	AAAATATACA	CAAATCTATT	ATTAAAAATT	19260
AAAATTTATC	AAAACATAAT	GCATACAAAC	TCTTAGACTA	TACATGGCAC	CATTCATAGT	19320
CAACAGAAAT	GTAAACAAAC	ATAAAGATGC	AATATTGTCA	TAACTGCATA	AAATATAGCA	19380
CATAATGTGC	TAGTATAATA	ATTTTGCAGT	CACCTCTTGT	TGGTATTGCA	GTGAGCTCAA	19440
GTGTTTTGAG	TATCTACTTA	AAATGCTGTG	TGACATTAGT	CATTTTCACC	TGAGCAGTTC	19500
ATATCTCCAG	TAAATTCTGC	CTCACAGTAA	AAAGTGATCT	CTCAAGGTTT	TCACATATTT	19560
TTATCATGTT	TAGTGCAATA	CCTTAAGCCT	TTAATAACAC	CATGGGCTCC	ATATGAAGTG	19620
TCATTAATGA	TGTTGGAAGT	GCTCCCAAGA	AGCAGAGAAA	AGTTATGACA	TTATAATAAA	19680
AAAATTGAGT	TGCTTAATGT	ATACTATACA	TTGAGGTCTG	CAGCTATAGT	TGCCCACCAT	19740
TTCAAGATAA	ATGAATCCAG	TGCAACTATG	CCAGCAGGCA	TGAAATCTTG	CACTTTTTGT	19800
AAAATATCTT	TTTATTTTGG	ATTGAAAATG	CAGCTTTTTA	TGTGGGTGCA	GGATTGCTAT	19860

AAGGAAGTAT	ACATATAGAC	TCTAATATAA	TTTGAGAAAA	AGTGAAGTTA	TTATATGACA	19920
AAGCAAAAGG	AAGGTGAAGG	ATCTAGAGCT	GGAAAAGTTA	ATGCCAGCAA	AGGATGATTT	19980
GATTACATCA	GAAAGAGTTT	GGCTTCAAAA	ATGTCAAGAT	AACAGGAGAC	ACGCATGCTG	20040
CCAACCAAGA	AGAAGGAGAT	GAATTCCCAG	ATGTCATTAT	GAAAATCATT	GAGGAGAAAG	20100
GATTTCTGCC	TGAACAGATT	TTTAACACAG	ACAAAAGTGC	CCTATTCTGG	AAAAAAAAAA	20160
AAAAAAAAAG	CCACAAAGGC	CATTTATTAA	TAAGGAGCAG	AAGTGAGCAC	CAAGATTAGG	20220
CAGGAAAGAA	TAAGCTAACT	ACTGTTTTGT	GCAAATGCAG	TCAGATTTAT	GATCAGGATG	20280
GCCCCTACCT	ATGAAGCTAC	CCCCTCAAAC	CTTGAAGGGA	AAAGATGAAT	ATCAGCTTCC	20340
TATCTTTTGG	TTATACAAGA	CCCTTTTTCT	GGATTAGCTC	TGTCAATGCT	TTGTCCCTGA	20400
AGTCAGAAAG	TCCTTGCCAA	TAAGAGACTG	CCTTTTAAAG	TTTTTTTGAT	ATAGACAATG	20460
CCCCTGACCA	CCCAGAACCC	CATGAGTTCA	ACATGGAAGG	CATCGAAGTA	GTCTAATTTC	20520
CCCCAAACAC	AACATTCTAA	TTCAGCCTTT	ATATCAGGGA	GTCATAAGGA	CCTTTAAGGC	20580
TCATCACATA	CCATACTCTA	TGGAAAAGAT	AGTCAATGCT	GTGGAAGATA	ACCCAACAGA	20640
GAGAACATCA	TGAAAGTCTG	GAAGGATTAT	ACCATTGAAG	ATGCCCTAAT	TGTTATAGAA	20700
AAAGCCATGA	AAGCCATCAA	TCCTAAAACA	ACATATTTCT	CCTGGAGAAA	ACTATGTCCA	20760
GATGTTATAT	ATGACTTCAG	AGGATTTACA	ACAGACCAGT	CACAGAAATC	ATGAAAAAGA	20820
TTATGGATAT	GGCAAAATAA	AAAGGTGAGG	GTGAAGGGTT	TCAAGATATG	GATCATGGAG	20880
AAATTCAACA	GCTAATAGAC	ACCACTAATA	GACACTTTTA	ATTCCACACT	AGAGGAACTA	20940
AAAGATGACT	TGATGGAGAT	GAGTCCTTCC	AAAGCAGTGC	CAGATGAGAA	CGAAGACATA	21000
GAAAAAGCCA	TGCCAGAAAT	AAATTGACAT	TAGATCATCT	GGCAGACAGG	TTCCAGTTAT	21060
TTAAGACTTC	TTTTGACTTC	TTTTATATAA	CATGGACCCT	TCTATGATAC	AGGCACTGAA	21120
ACTAAAGCAA	ATGATAGAGG	AAGGATTACT	ACTATATAGA	AAATTTTATG	AGAAATAAAA	21180
AAGCAAAAGT	AGACAGAAAT	TATAATATAT	TTCCATAATT	ACACCAATGG	GCCTGCCTCT	21240
CCTGCCCCCA	ATTCTACCTC	CTCCATCTCT	TCCGCTTCTG	CCAGGCCTGA	AACAGCAAGA	21300
CCAACCCCTT	CTGTTTCTCC	TCCTACTCCT	CAGCCTACTC	AACATAAAGA	TGATAAGGAT	21360
GAAGACATTT	ATGATAAACC	ACTTCCACTT	AATGAATAGT	AAACATATTT	TTTCTTCCTC	21420
ATAATTTTCT	TAATAACATT	TTCTTTTCTC	CATATTACTT	TATTGTAAGA	ATAGTATTTA	21480

ATACGTATGA	CATATAAAAT	ATGTGTTAAT	CAACTGTTTG	CATTATTGGT	AAGGCTTCCA	21540
GTCAACAGGT	TATTAAGAGT	TAAGTATTTG	GGGAGTCAAA	AGTTATACAT	GGATTTTTGA	21600
CTGCAAGAGG	GCTCAATGCC	CCTAACCCCT	CAGTTCTTCC	AAAGTCAACT	GATATAGGAA	21660
GTTTCTTTAC	TTTTTCAAGC	ATTTAACATT	GCATTGATAT	GTCAACCTAA	AGGAAAACAC	21720
TGAGGCAAAT	TTAATATAAA	AAGAGAATTG	ATTTGGGCCA	AGTTTGAGGA	CTGCAACCCA	21780
GGAGCACAGA	GTCAAATTGC	CCTGAATATG	CACTCCGTTG	GCAGCAGTTA	CAAGTAGGTT	21840
TTTTTAAAGGA	AATACAAAAG	AGTCAACTTC	TAAGTTGTTT	ACCAAGAACT	TACATTAAAA	21900
TTATATAAGC	TATTGATTGG	CTATATACTG	TTCTTCGTAT	CACAAATCTT	ACGAACATGA	21960
AGATGATGAG	TGAGACAGCT	AGTCAGGAAT	AAAAATGCCT	TTTAACAATT	GCCGCCAGGC	22020
TTGGTAGAGG	GCAGCATGAC	AAGTCCCATA	CACGTGGCTC	TCTCAGCTTG	ATAAATTTTG	22080
CATACCTCAC	ATAGTGCAAA	CTACTCTGAG	CTATTTTTCT	TCTCTCACAT	TGAATGCCAC	22140
AATGTAGTCA	CCCATTGAGG	GCCTAGAGAA	GAAAAGAAAT	GGAACCTCA	GATTCAACAA	22200
AACCTCTCCT	GCACAACCTC	AGCCAGTTGA	CGAACAACTT	GCAGAGTTGG	GCACTTTTAT	22260
GTGCTAACAA	TTCATGCAGC	TTGATACCCT	TTCTTTTAGA	GCCCAGTAGA	AATAAAAATG	22320
AGGAAATAGA	GAGGTTAAAA	TGTTTCATCTT	ATTGCTTAAA	TGATAAGCTG	CTCTTCAGAG	22380
TTTCAAAAAG	CAAATTACAC	CATATTCCAA	CTAAAAGAAC	TATAGAGGCG	GAAAGGAGGT	22440
GATCTCTTTT	CTCTCTGTCA	TAAAAGGTAA	TGGCCAACAC	CCCTATAACA	AAAGACAGGT	22500
TAACAAGAGA	AAACGTGACA	GATTTATTAC	GTGCACATGT	GTGCATGAGA	GCCTTACAAA	22560
ACATGAACTC	AAAGGAGGGC	CAGATCATTC	ATGTTTAAAT	ATTCTCTTCA	CTGGGGTTAG	22620
GGGAGATGGA	AGTGTAAGG	TAAATGATTT	TTCAGAGGAA	ATTAATAAGT	CCAAAGAACA	22680
CAGATTAGAC	CAAGTTTCTC	TGGGCTTTGG	GGGAGGTGTA	ATCACCCAAC	AGATTCATCT	22740
TGCTCACTGC	CCAGAAAAGC	TGATGCCCTG	AGAACAGCAG	GTTTTTCCAA	TAGAGAGAGT	22800
TTAATAAACA	CACAGCTGTC	AGAGGCATTT	GAACCAGAGT	GACTCCATCT	TGAATAGGGG	22860
CTGGGTAAAA	TGAGGCTGAG	ACCCACTAGG	CTGCATTCCC	AGGAGGTTAG	GCATTCTTAG	22920
TCACAGGATG	AGAAAAGAGG	CCAGCACAAAG	ATTCAGGTCA	CAAAGACCTG	GCTGATAAAA	22980
CAGGATTCAG	TAACGAAGCT	GGCCAAAACC	CACCAAAAACC	AAGATGATGA	AAAAAGTGAC	23040
CTCTGGTCGT	CCTCACTACT	CATTATATGC	TGATTATAAT	GAATTAGCAT	GCTAGAAGAC	23100

ACTCCCACCA	GCACTGTGAT	AGTTTACAAA	TGCCATGGCA	ATATCAGGAA	GTTACCCTAT	23160
ATAGTCTAAA	AAGGGGAGGA	ACCCTCAGTT	CTGGGAACTG	CCCATTTCTT	TCCTGTAAAA	23220
CTTATGAATA	ATCCACCCCT	TGTTTAACAT	GTAATCAAGA	AGTAACTATA	AGTATACTCA	23280
GTTGAGCAGC	CCATGCCACT	GCTCTGCCTA	TGAAGTAGCC	ATTCTTTTAT	TCCTTTACTT	23340
TCTTAATAAA	CTCGCTTTCA	CTTTATGGAC	TGGCCCTGAA	TTCTTTCTTG	TGAGAGGTCC	23400
AAGAACCCTC	TTTTGGGGTC	TGGCCAGCTA	AACGGAAGGA	CAGGAGTTTA	TTACTACTCA	23460
AATCAGCCTC	CATGAAAATT	CAGAGGCTAG	ATTTTTTTAA	GGATAGTTTG	GTAGTCAGGG	23520
GCTAGGGAAT	GGGGAATGCT	GATTGGTTGG	GTCAGGGATG	AAACCATAGG	GAGTCAAAGC	23580
TTGTCTTCTG	GTCTTCCTGG	GAGGAGACCA	CATGACAAGA	TGAACCAGTT	TACCAGTCTG	23640
GGTAGTGCCA	GCCGGCCCAT	CAGAATGCAG	GGTCTGAAAA	ATATCTTGAG	CACCAATGGT	23700
AGGTTTTATA	ATGGTGATGT	TATCCATAGG	AGCAATTGGG	GACTTCTGAC	TGCATGACTC	23760
CTGAGCCCTA	ATTTCTTATC	TTGTGGCTAA	TTTGTTAGTT	CTACAAAAGC	AGTCTGATCT	23820
CCAAGCAAGG	AGGGGGTTTG	TTTTGGGAAA	GGGCTGTTAC	CATCTTTGTT	TCAAAGTTAA	23880
ACTGTAAACT	AAATGTCTCC	CATAGTTAGC	TTGGCCTATG	CTCAGGAATG	AATAATGGCA	23940
GCTTGGAGAT	TAGAAGAAAG	ATGGAGTAAT	TACATTTTTT	TTTCACATTT	TTTTCACTGT	24000
CACAATTTTT	TTAAAGGTGA	TTTCAGAGGT	AACATCACAG	GACATGGGAG	ACTAAAGGGA	24060
GGAAAGTATG	TCAAACAAAG	GCTGTCCTGT	TCTGCAGACG	AAACCTCACA	GAAAGCAACT	24120
CTCAGAGTCA	GTAGCCTATG	ATGAAAGTTT	CTCTGTCAGA	CATTCAGCAG	TGCCTGACTC	24180
TCAGTCTCTC	TCTCCTGCAA	GTTAATCTTT	CCTAGAGTGG	GCAAGGGAGG	CCTCCGAGAA	24240
AGCCTAGTTT	CCATCTTCTG	TTTACTTCCT	TTTATTTTCT	CCACAGATAA	AAATCTCCTT	24300
CACAAAAGGC	AGCTTTTCAG	GGCTGTTTCT	GTCTGCAGGC	CCTCTGAATA	GCCATCTCAA	24360
AATCTGTCAA	CGAAGTGTAT	ATTTTGCAGT	AAAATATTTT	TTGTTTTCTT	TAGTATGAAA	24420
CAATTTATAT	TATTAGATTA	CAGGAGTATT	AAAACCATCC	ATGATCTCAC	TTTTAAACAA	24480
ACCAATCTGA	AAGTCTAACA	TTGGGGCAGA	TTCTAAGCAA	TGTCTTATAA	AGAATAATTA	24540
TGTGTTAATG	AGTAAACTAA	GTTAATTAGT	CTCCTTAAAC	CAGAGGGTCA	GTTTACTCCA	24600
GGCCACATGG	TCAAAGGCAA	AAGTCCAACA	TTACATCAAA	CTCAAATAGA	GATTAGGAAG	24660
GAGGAGAAAA	GCAGCTCACT	TAGCTAAAGA	AAAAACAATA	AATTCAATTT	TGTGGAAAAG	24720

GAGGGCATAA ATGGAGGTGC TATCTAAAT GTTATTTTTC TGAAAGAAA AATAAGAAAT	24780
TAATGCTCCT ATTTGCAACT GTAACACTTA TTCCAGTATG TTCTCTTCTT TCTTCATGTT	24840
TGGCCAGAGC CAGACTTTGC AGACAGATTA CATCACATAT GTGGATGAGC TGGGCTCTTT	24900
CATAGGGGCC AAGCCTAACA TACCATGGCT CTTCTGACA GATCCCCGCC TGGCCCTGGA	24960
GGTGTACTTT GGCCCTTGCA GCCCATACCA GTTTCGACTG ATGGGACCAG GGAAGTGGGA	25020
TGGGGCCAGA AATGCCATCC TGACCCAGTG GAACCGGACA GTGAAGCCAA CCAGGACAAG	25080
AGTTGTCAGT GAAGTTCAGC GACCCCATCC CTTTTACAAT TTGCTTAAAA TGCTTTCATT	25140
CCCATTACTC CTTCTGGCTG TTACACTTAC ATTTTATTAA TGAGAAAGTC TTTGAGGTCT	25200
CAAAATTCAG CATAGAAGTG TAATCACACA ATACAACACA CACCACACAT ACACACACAC	25260
AATCACAACA TAGTTCCTCT CTCCTTTCCT GAAGATATGA AAATCAGTCT TGGCCCATTT	25320
GAATTAAAGT ATAAGTAAAA TGGAAAATAC TCAGCCTCTC TCTCTCTGTT GGGAATCTGT	25380
TCTCTAAAAG GCTTTTCACA TGCTGAATTG GCAAATTTGG GGATGCTTAA GATAAGACAG	25440
GAAGTTGAAT AAGCATGAGC ACAG	25464

(2) INFORMATIONS POUR LA SEQ ID NO: 5:

(i) CARACTERISTIQUES DE LA SEQUENCE:

- (A) LONGUEUR: 1605 paires de bases
- (B) TYPE: nucléotide
- (C) NOMBRE DE BRINS: simple
- (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADNc

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 5:

GCAATCATGA GCAAGAGGGT TGGCATCATC GGAGCTGGAG TCAGTGGCTT GGCTGCCATA	60
TGGTGCTGTC TGGAGGAGGG GCTGGAGCCC ACTTGCTTTG AAAGGAGCGA TGATGTTGGA	120
GGCCTGTGGA AATTCTCAGA CCACACAGAA GAAGGCAGAG CCAGCATTTA CCAGTCTGTA	180
TTCACAACT CTTCCAAAGA AATGATGTGC TTTCCAGACT TCCCTTATCC GGATGATTAC	240
CCAACTATA TACACCACAG CAAGCTCCAG GAATATATAA AGACATATGC TCAAAGAAG	300
GAACTTTTAA GATACATANA GTTTGAGACC CTGGTTTCCG GTATAAAGAA ATGTCCCAGC	360
TTCTTAGTCA CGGGCCAATG GGTTGTTGTT ACTGAAAGG ATGGGAAACA GGAATCTACT	420
ATTTTGTATG CTGTAATGAT TTGTTTCAGGA CATCACGTAT ACCCCAATCT GCCAACGGAT	480

TCCTTTCCTG GCCTGGACCA GTTTCGAGGC AACTACCTCC ATAGCCGGGA TTATAAGAAT	540
CCAGAAGCCT TCAAGGGGAA GAGGGTCCTC GTGATTGGTC TGGGGAATTC GGGATCTGAC	600
ATTGCTGTTG AGCTCAGCCG TCTGGCTACA CAGGTCATTA TCAGTACCAG AAGTGCTTCC	660
TGGGTCATGA GTCGGGTCTG GGATGATGGC TATCCTTGGG ATATGATGTA TGTTACCCGC	720
TTTGCATCCT TTCTCCGGAA TGTCCCTCCT TCATTCATCT CTGACTGGTT ATATGTCCAG	780
AAGATGAACA CGTGGTTTAA GCATGAGAAC TATGGCCTGA TGCCTTTAAA TGGTTCCCTG	840
AGAAAAGAGC CTGTGTTCAA TGATGAGCTC CCATCCCGCA TCCTGTGTGG CACTCTGTCC	900
ATCAAGCCCA GTGTGAAGGA GTTCACGGAA ACCTCAGCTG TGTTTGAGGA TGGGACCATG	960
TTTGAGGCTA TCGACTCTGT CATCTTTGCA ACAGGCTATG ATTATTCCTA CCCCTTCCTT	1020
GATGAGACCA TCATGAAAAG CAGAAACAAT GAGGTTACCT TGTTTAAAGG CATCTTCCCC	1080
CCACTAATGG AGAAGCCAAC CTTGGCTGTG ATTGGCTTGG TTCAGTCCCT TGGAGCTGCC	1140
ATCCCCACAG CAGACCTGCA AGCCTGGTGG GCTGCTAAAG TATTTGCAA CTCATGTACC	1200
CTGCCAACCA CGAATGAAAT GATGGATGAC ACTGATGAGA AAATGGGGAA AAAACTCAAG	1260
TGGTTTGGCC AGAGCCAGAC TTTGCAGACA GATTACATCA CATATGTGGA TGAGCTGGGC	1320
TCTTTCATAG GGGCCAAGCC TAACATACCA TGGCTCTTCC TGACAGATCC CCGCCTGGCC	1380
CTGGAGGTGT ACTTTGGCCC TTGCAGCCCA TACCAGTTTC GACTGATGGG ACCAGGGAAG	1440
TGGGATGGGG CCAGAAATGC CATCCTGACC CAGTGGAACC GGACAGTGAA GCCAACCAGG	1500
ACAAGAGTTG TCAGTGAAGT TCAGCGACCC CATCCCTTTT ACAATTTGCT TAAAATGCTT	1560
TCATTCCCAT TACTCCTTCT GGCTGTTACA CTTACATTTT ATTAA	1605

(2) INFORMATIONS POUR LA SEQ ID NO: 6:

(i) CARACTERISTIQUES DE LA SEQUENCE:

- (A) LONGUEUR: 532 acides aminés
- (B) TYPE: acide aminé
- (C) NOMBRE DE BRINS: simple
- (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: peptide

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 6:

Met	Ser	Lys	Arg	Val	Gly	Ile	Ile	Gly	Ala	Gly	Val	Ser	Gly	Leu	Ala	1	5	10	15
Ala	Ile	Trp	Cys	Cys	Leu	Glu	Glu	Gly	Leu	Glu	Pro	Thr	Cys	Phe	Glu	20	25	30	
Arg	Ser	Asp	Asp	Val	Gly	Gly	Leu	Trp	Lys	Phe	Ser	Asp	His	Thr	Glu	35	40	45	
Glu	Gly	Arg	Ala	Ser	Ile	Tyr	Gln	Ser	Val	Phe	Thr	Asn	Ser	Ser	Lys	50	55	60	
Glu	Met	Met	Cys	Phe	Pro	Asp	Phe	Pro	Tyr	Pro	Asp	Asp	Tyr	Pro	Asn	65	70	75	80
Tyr	Ile	His	His	Ser	Lys	Leu	Gln	Glu	Tyr	Ile	Lys	Thr	Tyr	Ala	Gln	85	90	95	
Lys	Lys	Glu	Leu	Leu	Arg	Tyr	Ile	Gln	Phe	Glu	Thr	Leu	Val	Ser	Gly	100	105	110	
Ile	Lys	Lys	Cys	Pro	Ser	Phe	Leu	Val	Thr	Gly	Gln	Trp	Val	Val	Val	115	120	125	
Thr	Glu	Lys	Asp	Gly	Lys	Gln	Glu	Ser	Thr	Ile	Phe	Asp	Ala	Val	Met	130	135	140	
Ile	Cys	Ser	Gly	His	His	Val	Tyr	Pro	Asn	Leu	Pro	Thr	Asp	Ser	Phe	145	150	155	160
Pro	Gly	Leu	Asp	Gln	Phe	Arg	Gly	Asn	Tyr	Leu	His	Ser	Arg	Asp	Tyr	165	170	175	
Lys	Asn	Pro	Glu	Ala	Phe	Lys	Gly	Lys	Arg	Val	Leu	Val	Ile	Gly	Leu	180	185	190	
Gly	Asn	Ser	Gly	Ser	Asp	Ile	Ala	Val	Glu	Leu	Ser	Arg	Leu	Ala	Thr	195	200	205	
Gln	Val	Ile	Ile	Ser	Thr	Arg	Ser	Ala	Ser	Trp	Val	Met	Ser	Arg	Val	210	215	220	
Trp	Asp	Asp	Gly	Tyr	Pro	Trp	Asp	Met	Met	Tyr	Val	Thr	Arg	Phe	Ala	225	230	235	240
Ser	Phe	Leu	Arg	Asn	Val	Leu	Pro	Ser	Phe	Ile	Ser	Asp	Trp	Leu	Tyr	245	250	255	
Val	Gln	Lys	Met	Asn	Thr	Trp	Phe	Lys	His	Glu	Asn	Tyr	Gly	Leu	Met	260	265	270	

Pro Leu Asn Gly Ser Leu Arg Lys Glu Pro Val Phe Asn Asp Glu Leu
 275 280 285
 Pro Ser Arg Ile Leu Cys Gly Thr Leu Ser Ile Lys Pro Ser Val Lys
 290 295 300
 Glu Phe Thr Glu Thr Ser Ala Val Phe Glu Asp Gly Thr Met Phe Glu
 305 310 315 320
 Ala Ile Asp Ser Val Ile Phe Ala Thr Gly Tyr Asp Tyr Ser Tyr Pro
 325 330 335
 Phe Leu Asp Glu Thr Ile Met Lys Ser Arg Asn Asn Glu Val Thr Leu
 340 345 350
 Phe Lys Gly Ile Phe Pro Pro Leu Met Glu Lys Pro Thr Leu Ala Val
 355 360 365
 Ile Gly Leu Val Gln Ser Leu Gly Ala Ala Ile Pro Thr Ala Asp Leu
 370 375 380
 Gln Ala Trp Trp Ala Ala Lys Val Phe Ala Asn Ser Cys Thr Leu Pro
 385 390 395 400
 Thr Thr Asn Glu Met Met Asp Asp Thr Asp Glu Lys Met Gly Lys Lys
 405 410 415
 Leu Lys Trp Phe Gly Gln Ser Gln Thr Leu Gln Thr Asp Tyr Ile Thr
 420 425 430
 Tyr Val Asp Glu Leu Gly Ser Phe Ile Gly Ala Lys Pro Asn Ile Pro
 435 440 445
 Trp Leu Phe Leu Thr Asp Pro Arg Leu Ala Leu Glu Val Tyr Phe Gly
 450 455 460
 Pro Cys Ser Pro Tyr Gln Phe Arg Leu Met Gly Pro Gly Lys Trp Asp
 465 470 475 480
 Gly Ala Arg Asn Ala Ile Leu Thr Gln Trp Asn Arg Thr Val Lys Pro
 485 490 495
 Thr Arg Thr Arg Val Val Ser Glu Val Gln Arg Pro His Pro Phe Tyr
 500 505 510
 Asn Leu Leu Lys Met Leu Ser Phe Pro Leu Leu Leu Leu Ala Val Thr
 515 520 525
 Leu Thr Phe Tyr
 530

(2) INFORMATIONS POUR LA SEQ ID NO: 7:

(i) CARACTERISTIQUES DE LA SEQUENCE:

- (A) LONGUEUR: 21 paires de bases
- (B) TYPE: nucléotide
- (C) NOMBRE DE BRINS: simple
- (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: Autre acide nucléique

- (A) DESCRIPTION: /desc = "oligonucleotide"

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 7:

TCACATAGAG TGCTATGGGG G

21

(2) INFORMATIONS POUR LA SEQ ID NO: 8:

(i) CARACTERISTIQUES DE LA SEQUENCE:

- (A) LONGUEUR: 25 paires de bases
- (B) TYPE: nucléotide
- (C) NOMBRE DE BRINS: simple
- (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: Autre acide nucléique

- (A) DESCRIPTION: /desc = "oligonucleotide"

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 8:

CTTAGGAAGA AGATAAAAAT GCAAC

25

(2) INFORMATIONS POUR LA SEQ ID NO: 9:

(i) CARACTERISTIQUES DE LA SEQUENCE:

- (A) LONGUEUR: 23 paires de bases
- (B) TYPE: nucléotide
- (C) NOMBRE DE BRINS: simple
- (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: Autre acide nucléique

- (A) DESCRIPTION: /desc = "oligonucleotide"

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 9:

AATGTCCATC ATCATAGTTC TCT

23

(2) INFORMATIONS POUR LA SEQ ID NO: 10:

(i) CARACTERISTIQUES DE LA SEQUENCE:

- (A) LONGUEUR: 23 paires de bases
- (B) TYPE: nucléotide
- (C) NOMBRE DE BRINS: simple
- (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: Autre acide nucléique

- (A) DESCRIPTION: /desc = "oligonucleotide"

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 10:

TAGGCTTGTG TAGCCTGCCC TCA

23

(2) INFORMATIONS POUR LA SEQ ID NO: 11:

(i) CARACTERISTIQUES DE LA SEQUENCE:

- (A) LONGUEUR: 16 paires de bases
- (B) TYPE: nucléotide
- (C) NOMBRE DE BRINS: simple
- (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: Autre acide nucléique

- (A) DESCRIPTION: /desc = "oligonucleotide"

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 11:

CCTCAGAGAG AACTAT

16

(2) INFORMATIONS POUR LA SEQ ID NO: 12:

(i) CARACTERISTIQUES DE LA SEQUENCE:

- (A) LONGUEUR: 16 paires de bases
- (B) TYPE: nucléotide
- (C) NOMBRE DE BRINS: simple
- (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: Autre acide nucléique

- (A) DESCRIPTION: /desc = "oligonucleotide"

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 12:

GGAGTCTCTC TTGATA

16

(2) INFORMATIONS POUR LA SEQ ID NO: 13:

(i) CARACTERISTIQUES DE LA SEQUENCE:

- (A) LONGUEUR: 16 paires de bases
- (B) TYPE: nucléotide
- (C) NOMBRE DE BRINS: simple
- (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: Autre acide nucléique

- (A) DESCRIPTION: /desc = "oligonucleotide"

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 13:

CCTCAAAGAG AACTAT

16

(2) INFORMATIONS POUR LA SEQ ID NO: 14:

(i) CARACTERISTIQUES DE LA SEQUENCE:

- (A) LONGUEUR: 16 paires de bases
- (B) TYPE: nucléotide
- (C) NOMBRE DE BRINS: simple
- (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: Autre acide nucléique

- (A) DESCRIPTION: /desc = "oligonucleotide"

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 14:

GGAGTTTCTC TTGATA

16

REVENDICATIONS

1. Séquence nucléotidique isolée, caractérisée en ce qu'elle est choisie parmi :
- 5 a) les séquences codant pour la protéine FMO2 ou FMOx humaine et leurs variants protéiques,
b) les séquences codant pour un fragment de l'une de ces protéines ayant au moins 10 bases,
c) les séquences génomiques FMO2 ou FMOx humaines et leurs
10 allèles,
d) les séquences présentant au moins 80 % d'homologie avec les séquences (a) et (c),
e) les fragments des séquences (c) ou (d) ayant au moins 10 bases,
15 f) les séquences qui s'hybrident avec une séquence de (a) à (e).
2. Séquence nucléotidique caractérisée en ce qu'elle est choisie parmi :
- 20 a) les séquences codant pour un polypeptide comprenant les amino-acides selon la séquence SEQ ID N° 3, ou selon la séquence SEQ ID N° 6,
b) les séquences nucléiques SEQ ID N° 1, SEQ ID N° 2, ou les séquences nucléotidiques représentées Figures 2 et 1,
25 les séquences nucléotidiques SEQ ID N° 4, SEQ ID N° 5, ou les séquences nucléotidiques représentées Figure 10, ou les séquences nucléotidiques codant pour les polypeptides correspondants,
c) un fragment d'une séquence selon (a) ou (b) comportant
30 au moins 10 bases,
d) une séquence qui comporte par rapport aux séquences (a), (b) ou (c) au moins une mutation ponctuelle, et
e) une séquence complémentaire des séquences (a), (b), (c) ou (d).

3. Séquence nucléotidique selon l'une des revendications 1 ou 2, caractérisée en ce qu'elle comprend au moins la mutation G.1263mac.A.
- 5 4. Séquence nucléotidique selon l'une des revendications 1 à 3, caractérisée en ce qu'elle comporte au moins 10 bases.
- 10 5. Séquence nucléotidique utilisable notamment comme amorce spécifique d'un allèle caractérisé en ce que sa séquence est choisie parmi les séquences selon la revendication 4.
- 15 6. Séquence nucléotidique utilisable notamment comme amorce nucléique caractérisée en ce que sa séquence est choisie parmi les séquences selon la revendication 4 et les séquences SEQ ID N° 7, SEQ ID N° 8, SEQ ID N° 9 et SEQ ID N° 10.
- 20 7. Séquence nucléotidique utilisable notamment comme sonde spécifique d'un allèle caractérisée en ce que sa séquence est choisie parmi les séquences selon la revendication 4 et les séquences SEQ ID N° 11, SEQ ID N° 12, SEQ ID N° 13 et SEQ ID N° 14.
- 25 8. Séquence nucléotidique selon la revendication 1, caractérisée en ce que la séquence code pour l'un des domaines de FMO.
- 30 9. Polypeptide codé par une séquence nucléotidique selon l'une des revendications 1 à 4, et 8, notamment les polypeptides de séquence SEQ ID N° 3 ou SEQ ID N° 6.
- 35 10. Vecteur de clonage et/ou d'expression dans une cellule hôte appropriée d'une séquence nucléotidique,

caractérisé en ce qu'il comporte une séquence selon l'une des revendications 1 à 3 et 8.

11. Vecteur selon la revendication 10, caractérisé en ce qu'il comporte les éléments permettant l'expression et/ou la sécrétion desdites séquences dans ladite cellule hôte.

12. Vecteur selon l'une des revendications 10 et 11, caractérisé en ce qu'il s'agit d'un vecteur à répllication autonome.

13. Vecteur selon l'une des revendications 10 et 11, caractérisé en ce qu'il s'agit d'un vecteur d'intégration chromosomique.

14. Vecteur selon l'une des revendications 10 à 13, caractérisé en ce qu'il s'agit d'un vecteur viral.

15. Vecteur selon la revendication 14, caractérisé en ce que le vecteur est réalisé sur la base d'un adénovirus, d'un AAV, d'un rétrovirus, d'un poxvirus ou d'un virus herpétique.

16. Cellule transformée par un vecteur selon l'une des revendications 10 à 15.

17. Cellule selon la revendication 16, caractérisée en ce qu'il s'agit d'une cellule procaryote.

18. Cellule selon la revendication 16, caractérisée en ce qu'il s'agit d'une cellule eucaryote.

19. Animal, caractérisé en ce qu'il contient une cellule selon l'une des revendications 16 à 18.

20. Procédé de production de polypeptide recombinant, caractérisé en ce qu'on cultive une cellule selon l'une des revendications 16 à 18 et en ce que l'on récupère la protéine produite.

5

21. Polypeptide susceptible d'être obtenu par la mise en oeuvre du procédé selon la revendication 20.

22. Polypeptide spécifique de la forme mutée d'un polypeptide selon la revendication 21, caractérisé en ce que sa séquence est choisie parmi les séquences polypeptidiques comprenant au moins une mutation.

23. Anticorps dirigés contre un polypeptide selon l'une des revendications 9, 21 et 22.

24. Anticorps polyclonaux ou monoclonaux selon la revendication 23, caractérisés en ce qu'ils sont obtenus par réaction immunologique d'un organisme humain ou animal avec un agent immunogène constitué par un polypeptide selon l'une des revendications 9, 21 et 22.

25. Anticorps polyclonaux ou monoclonaux selon la revendication 24, caractérisés en ce qu'ils sont obtenus par réaction immunologique d'un organisme humain ou animal à un agent immunogène constitué par un polypeptide selon la revendication 22.

26. Anticorps selon l'une des revendications 23 à 25, caractérisé en ce qu'il s'agit d'un anticorps marqué, notamment pour l'imagerie.

27. Utilisation de cellules selon l'une des revendications 16 à 18 ou d'un animal selon la revendication 19, pour la sélection de produits impliqués directement ou indirectement dans l'activité FMO.

28. Utilisation de cellules selon l'une des revendications 16 à 18 ou d'un animal selon la revendication 19, pour la sélection de produits interagissant avec la FMO naturelle ou mutée, notamment à titre d'agoniste ou d'antagoniste de cette enzyme.

29. Produit obtenu par la mise en oeuvre de l'une des revendications 27 ou 28.

30. Méthode de diagnostic d'une prédisposition à des troubles liés à FMO chez un patient, caractérisée en ce qu'on détermine à partir d'un prélèvement biologique dudit patient la présence d'une mutation dans le gène FMO par l'analyse de tout ou partie d'une séquence nucléique correspondant audit gène, la présence d'au moins une telle mutation étant indicative d'une prédisposition dudit patient à des troubles liés à FMO.

31. Méthode de diagnostic selon la revendication 30, caractérisée en ce que la mutation que l'on cherche à déterminer est la mutation G.1263mac.A.

32. Méthode selon l'une des revendications 30 ou 31, dans laquelle la séquence d'acide nucléique analysée est un ADN génomique, un ADNc ou un ARNm.

33. Méthode selon l'une des revendications 30 à 32, caractérisée en ce que ladite analyse est réalisée par hybridation.

34. Méthode selon l'une des revendications 30 à 33, caractérisée en ce que la présence d'une mutation est détectée par comparaison avec la séquence correspondante naturelle non mutée.

35. Méthode selon l'une des revendications 33 et 34, caractérisée en ce que ladite hybridation est réalisée à l'aide d'au moins une sonde oligonucléotidique spécifique de l'allèle.

5

36. Méthode selon l'une des revendications 30 à 35, caractérisée en ce que ladite analyse est réalisée par séquençage.

10

37. Méthode selon l'une des revendications 30 à 35, caractérisée en ce que ladite analyse est réalisée par migration électrophorétique, et plus particulièrement par SSCP ou DGGE.

15

38. Méthode selon l'une des revendications 30 à 35, caractérisée en ce que ladite analyse est réalisée par une méthodologie visant à détecter une troncation de la protéine.

20

39. Méthode selon l'une des revendications 30 à 38, caractérisée en ce que tout ou partie de la séquence nucléique du gène FMO est amplifiée préalablement à la mise en évidence de la ou des mutations.

25

40. Méthode selon la revendication 39, caractérisée en ce que l'amplification est réalisée par PCR ou PCR-like.

30

41. Méthode selon l'une des revendications 39 ou 40, caractérisée en ce que les amorces choisies pour réaliser l'amplification sont choisies parmi les amorces définies selon l'une des revendications 5 ou 6.

35

42. Réactif pour détecter et/ou identifier une mutation du gène FMO dans un échantillon biologique, caractérisé en ce qu'il comprend une sonde dite de capture et/ou une sonde dite de détection, l'une au moins de ces

sondes comportant une séquence selon l'une des revendications 1 à 8, ou en ce qu'il comporte un anticorps selon l'une des revendications 23 à 26.

5 43. Méthode de diagnostic d'une prédisposition à des troubles liés à FMO chez un patient, caractérisée en ce qu'on détermine à partir d'un prélèvement biologique dudit patient la présence d'une FMO mutée.

10 44. Méthode selon la revendication 44, caractérisée en ce qu'elle utilise un anticorps mono ou polyclonal selon l'une des revendications 23 à 26.

45. Méthode selon l'une des revendications 43 ou 44,
15 caractérisée en ce que la détection est effectuée par un procédé ELISA ou RIA.

46. Méthode de diagnostic d'une prédisposition à des troubles liés à FMO chez un patient, caractérisée en ce
20 qu'on détermine à partir d'un prélèvement biologique dudit patient l'activité enzymatique d'au moins une FMO.

47. Composition thérapeutique caractérisée en ce qu'elle comporte à titre de principe actif au moins un
25 composé capable de moduler l'activité FMO.

48. Composition thérapeutique selon la revendication 47, caractérisée en ce que le principe actif est capable de moduler l'activité FMO2 et/ou FMOx.

30 49. Composition thérapeutique caractérisée en ce qu'elle comporte à titre de principe actif au moins un composé capable d'interagir avec FMO.

50. Composition thérapeutique selon la revendication 49, caractérisée en ce que le principe actif est capable d'interagir avec FMO2 et/ou FMOx.

5 51. Composition thérapeutique selon l'une des revendications 47 à 50, caractérisée en ce qu'elle présente une activité différente sur FMO normale et FMO pathologique.

10 52. Composition thérapeutique selon l'une des revendications 47 à 51, caractérisée en ce qu'elle comporte à titre de principe actif un composé à activité pro-FMO.

53. Composition selon la revendication 52, caractérisée en ce que le composé à activité pro-FMO est choisi parmi les composés suivants :

- 15 a) une protéine ou un polypeptide selon la revendication 29,
b) un vecteur d'expression selon l'une des revendications
20 10 à 15,
c) une séquence nucléotidique selon l'une des revendications 1 à 4, caractérisée en ce que ladite séquence est une séquence sens induisant l'expression de FMO.

25 54. Composition thérapeutique selon l'une des revendications 47 à 51, caractérisée en ce qu'elle comporte à titre de principe actif un composé à activité anti-FMO.

30 55. Composition selon la revendication 54, caractérisée en ce que le principe actif est choisi parmi les composés suivants :

- a) un anticorps anti-FMO, selon l'une des revendications 23 à 26,
35 b) un vecteur d'expression selon l'une des revendications 10 à 15,

c) une séquence nucléotidique selon l'une des revendications 1 à 4, caractérisée en ce que ladite séquence est une séquence antisens inhibant l'expression de FMO,

5 d) une séquence nucléotidique selon l'une des revendications 1 à 4, caractérisée en ce que ladite séquence est une séquence sens inhibant l'expression de FMO.

10 56. Composition selon la revendication 49, caractérisée en ce que le principe actif est une séquence soluble interagissant avec FMO.

15 57. Utilisation d'un principe actif capable de moduler l'activité FMO, l'activité FMO2 et/ou FMOx, pour réaliser un médicament destiné au traitement et/ou à la prévention de troubles liés à FMO.

20 58. Utilisation d'un principe actif capable d'interagir avec FMO, avec FMO2 et/ou FMOx, pour réaliser un médicament destiné au traitement et/ou à la prévention de troubles liés à FMO.

25 59. Utilisation d'au moins un produit selon la revendication 29, pour réaliser un médicament destiné au traitement et/ou à la prévention de troubles liés à FMO, à FMO2 et/ou à FMOx.

30 60. Procédé de biodégradation ou de biosynthèse de composé organique ou inorganique, caractérisé en ce qu'il met en oeuvre un polypeptide selon l'une des revendications 9, 21 et 22 ou une cellule selon l'une des revendications 16 à 18.

35 61. Procédé d'élaboration de composé d'intérêt, caractérisé en ce qu'il utilise un polypeptide selon l'une

des revendications 9, 21 et 22 ou une cellule selon l'une des revendications 16 à 18.

62. Produit susceptible d'être obtenu par un procédé
5 selon l'une des revendications 60 et 61.

63. Utilisation de polypeptide selon l'une des revendications 9, 21 et 22, ou d'une cellule selon l'une des revendications 16 à 18 pour la détoxification de
10 composé xénobiotique.

1/53

FM02prot

MetAlaLysLysValAlaValIleGlyAlaGlyValSerGlyLeuIleSerLeuLysCys ATGGCAAAGAAGGTAGCTGTGATTGGAGCTGGGGTCAGTGGCCTAATTTCTCTGAAGTGC	60
CysValAspGluGlyLeuGluProThrCysPheGluArgThrGluAspIleGlyGlyVal TGTGTGGATGAGGGACTTGAGCCCACTTGCTTTGAGAGAACTGAAGATATTGGAGGAGTG	120
TrpArgPheLysGluAsnValGluAspGlyArgAlaSerIleTyrGlnSerValValThr TGGAGGTTCAAAGAGAATGTGGAAGATGGCCGAGCAAGTATCTATCAATCTGCTGTTACC	180
AsnThrSerLysGluMetSerCysPheSerAspPheProMetProGluAspPheProAsn AACACCAGCAAAGAAATGTCCTGTTTCAGTGACTTTCCAATGCCTGAAGATTTTCCAAAC	240
PheLeuHisAsnSerLysLeuLeuGluTyrPheArgIlePheAlaLysLysPheAspLeu TTCTTGCAATAATCTAAACTTCTGGAATATTTTCAGGATTTTGTCTAAAAAATTTGATCTG	300
LeuLysTyrIleGlnPheGlnThrThrValLeuSerValArgLysCysProAspPheSer CTAAAAATATATTTCAGTTCCAGACAACTGTCTTAGTGTGAGAAAATGTCAGATTTCTCA	360
SerSerGlyGlnTrpLysValValThrGlnSerAsnGlyLysGluGlnSerAlaValPhe TCCTCTGGCCAATGGAAGTTGTCACTCAGAGCAACGGCAAGGAGCAGAGTGCTGTCTTT	420
AspAlaValMetValCysSerGlyHisHisIleLeuProHisIleProLeuLysSerPhe GACGCAGTTATGGTTTGCAGTGGCCACCACATTCTACCTCATATCCCACTGAAGTCATTT	480
ProGlyMetGluArgPheLysGlyGlnTyrPheHisSerArgGlnTyrLysHisProAsp CCAGGTATGGAGAGGTTCAAAGGCCAATATTTCCATAGCCGCCAATACAAGCATCCAGAT	540
GlySerGluGlyLysArgIleLeuValIleGlyMetGlyAsnSerGlySerAspIleAla GGATCTCAGGGAAAACGCATCCTGGTGATTGGAATGGGAAACTCGGGCTCAGATATTGCT	600
ValGluLeuSerLysAsnAlaAlaGlnValPheIleSerThrArgHisGlyThrTrpVal GTTGAGCTGAGTAAGAATGCTGCTCAGGTTTTATCAGCACCAAGGCATGGCACCTGGGTC	660
MetSerArgIleSerGluAspGlyTyrProTrpAspSerValPheHisThrArgPheArg ATGAGCCGTATCTCTGAAGATGGCTATCCTTGGGACTCAGTGTCCACACCCGGTTTTCGT	720
SerMetLeuArgAsnValLeuProArgThrAlaValLysTrpMetIleGluGlnGlnMet TCTATGCTCCGCAATGTACTGCCACGAACAGCTGTAAAATGGATGATAGAACACAGATG	780
AsnArgTrpPheAsnHisGluAsnTyrGlyLeuGluProGlnAsnLysTyrIleMetLys AATCGGTGGTTCAACCATGAAAATTATGGCCTTGAGCCTCAAACAAATACATTATGAAG	840

Figure 1

2/53

FMO2prot

GluProValLeuAsnAspAspValProSerArgLeuLeuCysGlyAlaIleLysValLys
 GAACCTGTACTAAATGATGATGTCCCAAGTCGTCTACTCTGTGGAGCCATCAAGGTGAAA 900
 SerThrValLysGluLeuThrGluThrSerAlaIlePheGluAspGlyThrValGluGlu
 TCTACAGTGAAAGAGCTCACAGAACTTCTGCCATCTTTGAGGATGGAACAGTGGAGGAG 960
 AsnIleAspValIleIlePheAlaThrGlyTyrSerPheSerPheProPheLeuGluAsp
 AACATTGATGTCATCATTTTGTCAACAGGATATAGTTCTCTTTCCCTTCCTTGAAGAT 1020
 SerLeuValLysValGluAsnAsnMetValSerLeuTyrLysTyrIlePheProAlaHis
 TCACTCGTTAAAGTAGAGAATAATATGGTCTCACTGTATAAATACATATTCCCCCTCAC 1080
 LeuAspLysSerThrLeuAlaCysIleGlyLeuIleGlnProLeuGlySerIlePhePro
 CTGGACAAGTCAACCTCGCGTGCATTGGTCTCATCCAGCCCTAGGTTCATTTCCTCCA 1140
 ThrAlaGluLeuGlnAlaArgTrpValThrArgValPheLysGlyLeuCysSerLeuPro
 ACTGCTGAACCTCAAGCTCGTTGGGTGACAAGAGTTTCAAAGGCTTGTGTAGCCTGCC 1200
 SerGluArgThrMetMetMetAspIleIleLysArgAsnGluLysArgIleAspLeuPhe
 TCAGAGAGAACTATGATGATGGACATTATCAAAAGGAATGAAAAAGAATTGACCTGTTT 1260
 GlyGluSerGlnSerGlnThrLeuGlnThrAsnTyrValAspTyrLeuAspGluLeuAla
 GGAGAAAGCCAGAGCCAGACGTTGCAGACCAATTATGTTGACTACTTGGACGAGCTCGCC 1320
 LeuGluIleGlyAlaLysProAspPheCysSerLeuLeuPheLysAspProLysLeuAla
 TTAGAGATAGGTGCGAAGCCAGATTTCTGCTCTCTCTTCTTCAAAGATCCTAACTGGCT 1380
 ValArgLeuTyrPheGlyProCysAsnSerTyrxxxTyrArgLeuValGlyProGlyGln
 GTGAGACTCTATTTTCGGACCTGCAACTCCTATNAGTATCGCCTGGTTGGGCCTGGGCAA 1440
 TrpGluGlyPheArgAsnAlaIlePheThrGlnLysGlnArgIleLeuLysProLeuLys
 TGGGAAGGATTCAGAAATGCCATCTTCACCCAGAAGCAAAGAATACTGAAGCCACTCAAG 1500
 ThrArgAlaLeuLysAspSerSerAsnPheSerValSerPheLeuLeuLysIleLeuGly
 ACTCGGGCCCTGAAGGATTCATCTAATTTCTCAGTTTCTTTCTGTTGAAAATCCTGGGC 1560
 LeuLeuAlaValValAlaPhePheCysGlnLeuGlnTrpSer***
 CTTCTTGCTGTTGTTGTGGCCTTTTTTGGCCAACTTCAATGGTCCTAG 1608

Figure 1 (suite)

3/53

fragment1.FMO2

```

CACAGAAGTT ATGGAAGTAT ATATNTNACA CAGCNAATGC AGATGGGGTTC ATACAAAAAA
  10          20          30          40          50          60
AAAATTGAGT CATAGGCGAG GGAAGAGAAA ACCAGTTTGG TTGTTATTTC TCTAAGGTAT
  70          80          90         100         110         120
AAGGCAACTT AATGAGAGAA CTTGTTGGTT TGTATTCTCT CTAAGGTGTA AGGCAACTTA
 130         140         150         160         170         180
ATGAGAGAAC TTGTTGGTTT GTTATTTCTC TAAGGTGTAA GGCAACTTAA TGAGAGAACT
 190         200         210         220         230         240
TGTCTCACA ACTCTGTGCT TCTCAAAAAG TACCATGTGG TTTTAAGGAG TTTTCTGTTG
 250         260         270         280         290         300
AAATATGTAA CAGAAAGTTA AAAATGAAGG AGGAGATACC AGGGCTGATA TGTATATTGT
 310         320         330         340         350         360
TCTTATGATA ATATACACAA TTAAGGGAGC TTGGACTCAT CCAAGAATGG TTCAGGGGGG
 370         380         390         400         410         420
TGAATATGGG GCCCCTAACC ATGTAGAGGG GTGAAGGGGG CCAACACCCT CTTATTCTCC
 430         440         450         460         470         480
TATAGTAACC GGAGTATTTT CAAATCCAAA TACTTCTCAG CATTTTTATG CCTATCTTAA
 490         500         510         520         530         540
AACTTTAATT GATGATAAGA TTTCACATCT CAGACCCTAG ACTCTTGCAG TTAGTGACAT
 550         560         570         580         590         600
TACTCCTCTT TCACATGCTG CCTGAATTCC TCTCTTATCC ACTAACAAGT GGTCATCTAA
 610         620         630         640         650         660
CCTTTGTTTT ATATTTACAG TGATAGGGAG TTCACTACTA TACAAAACAA CTCACTCTAT
 670         680         690         700         710         720
ATTGTGTACG CACGGTTGGC TAGAAAGTTC TTTCTCACAT TGAGACAAAC TCAGTTTTTT
 730         740         750         760         770         780
TGTATCATTA ACATATGTAA TCCTAGAGAG ACTTGTCTCT ACCCAAATA GGCTTGTATG
 790         800         810         820         830         840
ACCTTTTCTG TATAGGAACC TTCTTGAAGG CAATTGTCAT GGATTTTTC AATCTCCTC
 850         860         870         880         890         900
TAGGCTAAGC ATCAGAAGTT TCTTCAGTGG CCCCTTATGT GACATGGCTT TAAGAAGTTC
 910         920         930         940         950         960
ATTGCCATCT TGTTATTCTT TAAATTCTCT GGTAGCTAAG CCAAGCTTAT TTCCACATGC
 970         980         990        1000        1010        1020
GTGAAAGATG GATGGGTGGT AAGGCCTATG GTCCTGAACT AGTAAATGAA TTAACATATG
 1030        1040        1050        1060        1070        1080
TGAAAGTATT TTTACAACCT AAAACACTTT GTAAACACAA GATACTATTA AAACATATCT
 1090        1100        1110        1120        1130        1140
CCTTCATAGA AGGCAACTTG GCAAAGCATA TTAGAACTT GCAACTCAAA ATATGGTCTA
 1150        1160        1170        1180        1190        1200
AGGGTACACC TGGCATCACC TGGAAGCTTG CAAGAAACAA TTTCAGGCCC ACCTCATATC
 1210        1220        1230        1240        1250        1260
CATTAAATCC CACCTGCATT TTGACAAGAT CTCCAAGTGT TTCATTTGCA CATTAAAGTC
 1270        1280        1290        1300        1310        1320
TGAGGAGCTC TGTAACCTTA AAAATATGCA AACACTTTCA CTAAGCAATC CTTGTTTTTA
 1330        1340        1350        1360        1370        1380
GGTATTTATC CTGAGAAAAT GATTAGGTGA AGTACACAGC ATTTTATATA CAGCAAAATC
 1390        1400        1410        1420        1430        1440

```

Figure 2

4/53

fragment1.FMO2

TTACAAAGAA	AGTTCCACTA	ATAGAAAATT	GGCCAAGGTG	CTTTACGGTT	CTTCTATGGA
1450	1460	1470	1480	1490	1500
TACACCATT	GCCAGCCAAG	GCTAACTCTA	TAAAGCAAGC	TGGTTCTTCT	CAGCTGGTGC
1510	1520	1530	1540	1550	1560
TGGTACCTAG	AAGATGCCTG	CTCCTCAGCT	GATGCATGCC	ACCATGAACA	CAGATGTATG
1570	1580	1590	1600	1610	1620
TTTAGTTGAC	ACAAAAGGCC	ATTAATATCC	CCACATCAGT	TATCCCTGGA	GGATAACTAA
1630	1640	1650	1660	1670	1680
GCCATCTGCC	TCCATCATCT	TTTAAGGGTT	CAGTCAGTTT	AAACTTTTGC	TTCTATACCT
1690	1700	1710	1720	1730	1740
AGGTATTTTC	TTTCTCTGT	ATGTTGGTCA	GGTACAATTA	TTTTTAACAG	GGCTTCCATC
1750	1760	1770	1780	1790	1800
AATATCATAA	CTACCTAGAG	AAGACATTGC	AAAGATAAAA	TTGGAGAATT	GTTAACAGGC
1810	1820	1830	1840	1850	1860
TGTTAACAAA	ATGTGTACCC	AACTGCCAAT	GAAGTGGCTT	GATTTTTTTC	TTTTTTTAAA
1870	1880	1890	1900	1910	1920
ATTTTTCTTT	TGTATCCTTT	TATTTTATTT	ACTTATTTTT	TAGAGACACA	GTCTCGCTCT
1930	1940	1950	1960	1970	1980
GTTACCCAGG	CTGGCGTACA	ATGGCACAAT	CATAGCCCAC	TGCAGTCTCG	ACCTCCAGGG
1990	2000	2010	2020	2030	2040
CCTCAAGCAA	TCCTCCCACC	TCAGCCTTCC	AAGTAGCTGA	GACTACAAGT	GCATGCTGCC
2050	2060	2070	2080	2090	2100
ATGCCTGACT	GATTTTTTGT	TTTTTGCAGA	GATGAAGTCT	CACTATGTTG	CCCAGTCTGA
2110	2120	2130	2140	2150	2160
CCTTGAAGTC	CTAGCAATAC	CCTACCCTGG	CCTCCCAAAC	TGCTGGGATT	ACAGGCATGA
2170	2180	2190	2200	2210	2220
GCCACTGTGC	CTGGCCTTTT	TTCATTTTAA	CTGAGAAATG	TGTTTCAGCTC	TTTTGTTCCT
2230	2240	2250	2260	2270	2280
TAGTCATTGA	TCATCACTTT	TGTTATATCT	GTTAGTCTTG	TCATAGAGTT	GCTGCACTTA
2290	2300	2310	2320	2330	2340
TTACACAGAG	AAGGCCTTTT	ATCACGACCA	ATTTATTTTA	GGAAATTTCA	GGGAAAACGT
2350	2360	2370	2380	2390	2400
TTTTCTAGAA	CACCTTATTT	GACATTATAA	AACAACCTTT	CACTCTTGCA	CTCCAGACCT
2410	2420	2430	2440	2450	2460
CCCTTTCCAG	TTTTCTTTTT	CTCCATAGTG	GTCATCACCA	CTTGTTTTAT	TTTATTGATG
2470	2480	2490	2500	2510	2520
GGCTGTCTGG	CTCCCTCAAC	TGCAAAGTAA	ACTCCACAAA	GGCAGAGAGT	TTTGTCTCTT
2530	2540	2550	2560	2570	2580
TTATTCAATTG	CTGTACCTGC	ATCACTTAGA	AAGTTTCTGG	CACCTAGGAA	GTGTTCACTA
2590	2600	2610	2620	2630	2640
AATATTTATT	GAATAAGTTT	ATGTAAAACG	TCTCAGACTC	CTTAGAGAAA	CTGGTCTTTT
2650	2660	2670	2680	2690	2700
GGGGTTGGAG	AATAAAGTTC	TTTACCTCAT	CAGTTAGACT	CTATCTAAGG	TACACGAGGG
2710	2720	2730	2740	2750	2760
CTTGCTAGTC	TCCTAAGTTA	GTCTGCTAAT	AAATGTTAAC	CCTAATAACT	GAAATTATTA
2770	2780	2790	2800	2810	2820
GCAGAGGTAA	TTATCCAGTT	CTATATCAAG	GCAAAAAGAC	AGCAGTGGAT	AGAAAGATCT
2830	2840	2850	2860	2870	2880

Figure 2 (Suite)

5/53

fragment1.FMO2

TAGAAAGTCCC	ACTAGGTTCA	TCCAAGCCAC	CATACACATA	GGCAGAAAAA	TCAAAATAAG
2890	2900	2910	2920	2930	2940
ATATGAGCCT	GGACAGGGTG	AGCAATCTGG	GAAAAGATGA	ACACAGTATG	CTAGGACCCA
2950	2960	2970	2980	2990	3000
GAAATCATCA	AGTCTATGAA	AACTAAGCCA	GAACACAAAT	GTGAATTCCA	TAAGATCAGG
3010	3020	3030	3040	3050	3060
AACATAATCT	GTCTTGTTC	TCCAGGCATG	GTAATCTGCC	AGAAATAGTG	CTTAACTGCA
3070	3080	3090	3100	3110	3120
AGAACTGAAT	ATTTGTTAGA	TAATTAAACC	ATCAACTAAA	TGAGATTTCAT	GCAACCATGA
3130	3140	3150	3160	3170	3180
AAAATGCTGC	TATAGGTACA	CAATATTGAT	ATACTAGAAA	GTTAAAAAAT	CAAGTTGGAA
3190	3200	3210	3220	3230	3240
ATTAGACTAT	TCCATTTCTG	TTTGTGTGTA	TGTATCTACA	AATAGGTGGA	AGGATATACC
3250	3260	3270	3280	3290	3300
AAAATGTCAA	CAGCAGTTAC	CTCTGGGTGG	TGAGGAGTAA	TCTTAACCTT	GTTATTTATC
3310	3320	3330	3340	3350	3360
CCTATATGTT	CATTTGTGAA	TGAATATTTA	TTACATCATT	ATAAAAAGGA	TTTTTAAACT
3370	3380	3390	3400	3410	3420
ATCTGTATGT	TTAAGAGTAT	ATGTTGCTAC	TATGTAAGAG	TATATGCTGT	TACTGTAAAG
3430	3440	3450	3460	3470	3480
ACATTGCATT	ACTACTGTTG	ACCTCAGAGC	ACGCGCCTCT	TGCCTAATTC	TAGGACTCCT
3490	3500	3510	3520	3530	3540
AACTAAGTCT	TTGGAGTTTC	AGCTGGAAGA	ATGCTGGAGG	AATACGGAAC	TCCTCCCAT
3550	3560	3570	3580	3590	3600
TCTCACAGCC	ACCTCCAAC	CTTAAAAACG	CTTCCAAC	CCTCCCAGCA	CACAACCAAG
3610	3620	3630	3640	3650	3660
GGAGAAAAC	ATTCTGTCAA	AGAGACGGTG	CCAAAAGGCA	AAAACAAAGG	TAAGGATGAT
3670	3680	3690	3700	3710	3720
CGCTGGGGAA	AGAAGCTGAA	AAGGAAAAGC	TCAGAACTCT	AGCTGGAAAT	TTGGCTCACA
3730	3740	3750	3760	3770	3780
TCCCTAGTAT	GTTACTGCAT	AGTCTGGCTT	TGTTCAATGG	GTCGCTTTTA	AATATTAAAG
3790	3800	3810	3820	3830	3840
CTAGATGTAA	GCAAGGTTTG	CAACAAAGTC	CATAAGAAAC	TCAGCTTTTC	TCAAAGGCAA
3850	3860	3870	3880	3890	3900
GAAGAGAGCA	GGATTTTGA	CTGGCTCTTT	ATTCAATAGT	GCTGCTTATT	AAATTACCAC
3910	3920	3930	3940	3950	3960
TGCTACAATG	TTTAAAGCCA	ATTACCTGAG	CACATCATAA	GGATTCTCTT	ACCGGTTGTC
3970	3980	3990	4000	4010	4020
CCAGTTAAGT	AATGTTGATT	GATCAACTCC	TTGACAGGAG	CTGATGGCAA	AGAAGGTAGC
4030	4040	4050	4060	4070	4080
TGTGATTGGA	GCTGGGGTCA	GTGGCCTAAT	TTCTCTGAAG	TGCTGTGTGG	ATGAGGGACT
4090	4100	4110	4120	4130	4140
TGAGCCCACT	TGCTTTGAGA	GAAGTGAAGA	TATTGGAGGA	GTGTGGAGGT	TCAAAGTAAG
4150	4160	4170	4180	4190	4200
TGAGATTTTC	TTGGGTCTTG	AACAGGTTGT	GTTGTTATTT	CAGGGTGAAT	CACAGTTACT
4210	4220	4230	4240	4250	4260
GATGGGTCAT	ATTGAGAAAT	TTATTAAACA	ACTCTGATCA	GATTTTATTT	CTATTTATTG
4270	4280	4290	4300	4310	4320
ATGTGGCCAT	AATGGAAC	AAGTCATAGG	CTGGCATCTC	TCCCCAGTC	AATACTAACC
4330	4340	4350	4360	4370	4380
CAACCCAGGT	AGCTGACCCA	GGCATGTAAA	AGATCTCTTC	TTTTGGATT	AGCAATTGTC
4390	4400	4410	4420	4430	4440

Figure 2 (suite)

6/53

fragment1.FMO2

```

TTACAGCCCA TACTTCTGTC ATTCTTTAAT ACGCTAATAT TAGAGAACAT TTTACAAAAA
  4450      4460      4470      4480      4490      4500

TAGAAGTAAC AGGGATTCTT CTCAAGATAT CACTTCTGTT TCAATTATTA AACCAAATGC
  4510      4520      4530      4540      4550      4560

TTCTTTAGAG ACCATGCTCT TATCATTACT ATTTTCTCTT GACAAATGAA GCATGTTTGT
  4570      4580      4590      4600      4610      4620

TTACTGAGCT TTATCAATGA CATTCTAGTA TAACTGCTGT GAAACTCTTT GTTAAATATG
  4630      4640      4650      4660      4670      4680

TTTTATTAAA TTTATTCTAT TAATCAAACC AAAATATTGA TAATGCTATT TGTCTGTATT
  4690      4700      4710      4720      4730      4740

AGTCCATTCT CATGCTGCTA TGAAGAAATA CTGAGACTGG GTGATTATTA AAGGAAAGAG
  4750      4760      4770      4780      4790      4800

GTTTAATTGA CTCCCAGTTC CACAATGCTG GGGAGGACTC AGGAAATTTA CAATCATGGC
  4810      4820      4830      4840      4850      4860

AGTGGGAAAG AGAGGTGCTG AGCAAAGGGG GAAAAGCCCC TTATAAAACC ATCAGATCTC
  4870      4880      4890      4900      4910      4920

ATTAGAACGC ACTCACTATC ATGAGAACAG CATGAGGGTA GCTGCCCCCT TGATTCAATT
  4930      4940      4950      4960      4970      4980

ACCTACCCCC ACCAGGTCCT TCCCAAGACA TGTGGGGATT GTGGGAAC TAATTCAATA
  4990      5000      5010      5020      5030      5040

TGAGATTTGG ATGGGGACAC AAAGCTAAAC CATGTCACTG TCCTTAAAAA TTTGTATAAA
  5050      5060      5070      5080      5090      5100

ACTTAGAAAG TTGCATAGAT AGCTATAAGG AGTTACAATT ATTCCTTCCC ACAACCTCTC
  5110      5120      5130      5140      5150      5160

AATAGGTAGT AGCTTACCAC CTTCTAGCTG TGAGATCTTG AGCAAGTTAT TTACATCTCTG
  5170      5180      5190      5200      5210      5220

TGTTTCAATT TACTCAGTTA TAAATGGATA TAATAACAGG AAAGTGTGAT TATCTCATAG
  5230      5240      5250      5260      5270      5280

TGCTATTTTG AAGATTAAGG GAGATAATTC ATATAAAGAA CTTAGATAAG TTCCGGACTC
  5290      5300      5310      5320      5330      5340

ATAGAGTTCA ATAAATGTTA GCTACTAATA ATAACATAT ATTTTATAGA TGAGCAAAC T
  5350      5360      5370      5380      5390      5400

GAAAGTGAGG GAGGTTAAGT GAGATGGCCA GGGCCACACA ACTGGAGGAA CTGGCCTTCA
  5410      5420      5430      5440      5450      5460

AACCACGGCC TACGTGACTT CTAAACAGAT AAGCCCTGAC TTACAACCAT GCCCTAECTT
  5470      5480      5490      5500      5510      5520

GCATTCTTGC TCAAAAAGAT TAAACAAAAG TTTAAGTTCA GAACCCAAAA GCAATGACTT
  5530      5540      5550      5560      5570      5580

TAGAATTATG TAATCAGGTA TCCCTGAGAT ATTAAACAC ATAAGAATAT TCCAAATGGG
  5590      5600      5610      5620      5630      5640

AGCAAAAGGT TTGAATACAT GAAAATCAAA CTCATATCAG CAGAGACCAT ATAAAGGGCT
  5650      5660      5670      5680      5690      5700

CTCACTGCAG GCTGACTAGT TAGGAGGATG GCAAGGTGAT CCAGGACCTG CGCATGCTTT
  5710      5720      5730      5740      5750      5760

GTCAGTTCAA ATTGAATCTC ATGCCAACAG CGATCTTTT TAACATGTAA CATTAGGTGT
  5770      5780      5790      5800      5810      5820

CTCAGGTACA CATGACCATA AACCACACCT GGAGGGTTTC TTTTATTTTC TTTTAAATAT
  5830      5840      5850      5860      5870      5880

TTTTCTGAGA CAGGGTCCTA CTCTGTCACC CAGGCTACCA TGCCCAGCCA TGGAGAGTTT
  5890      5900      5910      5920      5930      5940

CTTAAAGATA CTGATTCCTT TGGTTAAACC TGCCACCAAN AAAAAAA
  5950      5960      5970      5980

```

Figure 2 (suite)

7/53

fragment2.FMQ2

```

ACAAAGACTG GCATTAAATA GGTAGAGACC TAGGATGCTA ATATCTTGCA ATGTGCCAAA
  10         20         30         40         50         60
ATAATTGTCC CTGTCCCCAA CCTCACCATT GCCAATATTA CCCCTACCCC TCACAGTGAG
  70         80         90        100        110        120
CGTCACAGGC AGGCAACAAA CTGGTGTCGT CACAGAATGA TTGATGGAAC ACATAGACTG
 130        140        150        160        170        180
CATTCAATTAC CTAAACATTG TCGTCACACT GCAGCAACCA AAGACAATCG CATTACCCAG
 190        200        210        220        230        240
GGGTTAGATG TAGGAAGAGT AAAAAACAAA AAATTTTGA ATGCGTAATT ATCACTAATT
 250        260        270        280        290        300
ATTTTATTTG ATCCTTCAGG AGAATGTGGA AGATGGCCGA GCAAGTATCT ATCAATCTGT
 310        320        330        340        350        360
CGTTACCAAC ACCAGCAAAG AAATGTCCTG TTTCACTGAC TTTCCAATGC CTGAAGATTT
 370        380        390        400        410        420
TCCAAACTTC CTGCATAATT CTAAACTTCT GGAATATTTC AGGATTTTTG CTAAAAAATT
 430        440        450        460        470        480
TGATCTGCTA AAATATATTC AGTTCAGGT ATTGTATTTT TGGGGAAATG GGTTTCTCTG
 490        500        510        520        530        540
CATTAGTTCA GCTCATATTT AGATAGAAAA GTTACTCTGA TAATGAAAGC AATTATGAAT
 550        560        570        580        590        600
GAAGTATCCC ATTCTAAGTA TTTGTTGAAA TATAACAGCC TCATATAAAA CCCCCAAAGT
 610        620        630        640        650        660
AGTGTTCATTA CCCTTGGTAT TATAGATTAT ATACATTAAT TGAAGAGGAA AATCATCTGT
 670        680        690        700        710        720
TAAAATTAAA GGGTTGAATA ATAATATATT GATGTCAAAA CTTTT
 730        740        750        760

```

Figure 2 (suite)

8/53

fragment3.FMO2

```

ccctgagaca gagtctcact ctgttgctca ggctggagtg cagtggcatg atctcagctc
   10         20         30         40         50         60
actgcaacct ctgccttcca ggctcaagtg attctcctgc ctcgccctcc agagtagctg
   70         80         90        100        110        120
ggattacagg cacacatcac catgcctggt tcatttatgt atttttggta gggacgaggt
  130        140        150        160        170        180
ttcaccattt gggccaggct ggtctcgaac tcctgacctc aagtgatcca cccgcctcgg
  190        200        210        220        230        240
cccccaaaag tgctgggatt acagggtgta accaccacac ccagcctcaa aaattcattt
  250        260        270        280        290        300
aaactaatat ctgttatcat tgaatacacc tagcttcatt tgccttgaaa gggcgtatac
  310        320        330        340        350        360
caaaattaaa ttgctgtttt gttttcttag cttcttcata gaaatgggat ttcttagatg
  370        380        390        400        410        420
tgtattaaat aaattcattg gtctctgttc atactagaag gctgtgggaa gtatttgctt
  430        440        450        460        470        480
atcatttttt tctgaatgca atctcttaca acctaaagat ggccagatca ttttgaaaaa
  490        500        510        520        530        540
cacttggaat taccttttcc tgtgcttccct caaaatcaac aaaaagcaat attttaatta
  550        560        570        580        590        600
agcatgctga atttttatca atggtctata ctttgagaaa tagctactat gcttagaaaa
  610        620        630        640        650        660
taaaatataa atcacatttc ttggccagggt atggtgattc atgtttgtaa tcccagcact
  670        680        690        700        710        720
ttggggaggct gaggcaggaa gatcacttga acccaagagt ctgagaccaa cctgggcaat
  730        740        750        760        770        780
acagtgaaaa tctgtctcta caaaaaattt ttaaaagatt atccaggcat gttgatcccc
  790        800        810        820        830        840
acctgtggtc ccagctattc tagactgaga agggaggatc gcttgagcct gggagggtcaa
  850        860        870        880        890        900
agctgcaata agtggtgatt gtgccactgc actccagcct gggcaacagt gtgagaccct
  910        920        930        940        950        960
gtctcaaaag aaataactaa catttcttga taaataactg ttagtgaggc ttatttttaa
  970        980        990        1000       1010       1020
tacatgtcat tttcttagta attctaatac taggcttata taatatcaac ttacaatagt
 1030       1040       1050       1060       1070       1080
aaattttggt gaaaatttgt atttataaat tccattaaaa tgtccagttc tacctaattg
 1090       1100       1110       1120       1130       1140
agtttttcac caattcctgg tagatctaac ttgtgaataa cagattatgt ataccagaag
 1150       1160       1170       1180       1190       1200
gttttgtaac tttgtgcact taactatcaa tctacttaac aaatatattg cttttttatg
 1210       1220       1230       1240       1250       1260
atatataact tctattccat tcttttaaag atcatgttag agtcgcaagg aagtcatttc
 1270       1280       1290       1300       1310       1320
tcttggttat tgtgttactg ctacttttgt ttcttgagaa gtgaagaggg gttgggaaga
 1330       1340       1350       1360       1370       1380
aagggttctg tttattggtc tctgagttgg tgtaagtcac aggtgtttaga gctcaactcg
 1390       1400       1410       1420       1430       1440

```

Figure 2 (suite)

fragment3.FMO2

```

agaagcaggg aaactgtaac aagccctggt gcttatgatt gtcaatgtaa tctacatcag
1450      1460      1470      1480      1490      1500

tgctttctcaa actttaatgt ggacatgaat cacctggata tcttggtaaa aatgttaggtt
1510      1520      1530      1540      1550      1560

ctaatttaat aggtatgggg taagtcttga aattctgcat ttctgacaag cttccaagtq
1570      1580      1590      1600      1610      1620

atactgaaga tcctgatcct caaatcacat ttggaatagc aaggatctac agcacttagt
1630      1640      1650      1660      1670      1680

taatatacta ctttgaacta ccatctgaaa tcttttcttt catctgaaaa ctgcccagat
1690      1700      1710      1720      1730      1740

atttaaagcc cttttacaag atttctacta atattccata tacattttta aattgagaca
1750      1760      1770      1780      1790      1800

gcttaaaaat taccaaccca gcagttggaa aaatatctga aaatttgaga tatataaaaq
1810      1820      1830      1840      1850      1860

actaaaatac ttgcaaatga gaagcatgcc attcctctag cattataaac ttgtcttcca
1870      1880      1890      1900      1910      1920

cttgacatcg tttcttaatc cagcagatat gaaacattta tgtacaattt taaaaattaa
1930      1940      1950      1960      1970      1980

cagacctcca gtgagctaca tttaaaaaaa tcaatgaacc aataaatcat tttattcaaa
1990      2000      2010      2020      2030      2040

taagatcatg aactgtcttg ctcacatgat gtactctggt ttaaaaatag caaatgttaa
2050      2060      2070      2080      2090      2100

aaactatcat tcagtggaa gctgaccatg tgtcaggcac tctgcaaagt gttttgcgtg
2110      2120      2130      2140      2150      2160

aaatatcttc tctaatacaa agtcacacaaa gaggcggcta cataaaacgt tcttgacata
2170      2180      2190      2200      2210      2220

tgccaattgc atgatcactt gaattattgg tttgtttcct tgttcagatt atcaaataac
2230      2240      2250      2260      2270      2280

aaacagagag aagttcttta aaagaaaaga tatatatattg gtgatagagc attgtaatga
2290      2300      2310      2320      2330      2340

gaatgtacat gccatggtaa actattttgtg tattcaggga gttaaaggaa gacaaagggtt
2350      2360      2370      2380      2390      2400

tttaaatggg gaaaaaatac aattacataa ttgttttgaa ataattatat aaagagcaat
2410      2420      2430      2440      2450      2460

aacaagggtg atgccagtct gagattggac agttactgag cagatgttct tgtagaagtc
2470      2480      2490      2500      2510      2520

atttttgtgt aagattatga tggctcttgt gtaagggtgt ggtttttgta gtttttgtaa
2530      2540      2550      2560      2570      2580

tcaggcacac atcatgagaa cccgctcttt ctggcctttt ccaattctat ttgtcgggtt
2590      2600      2610      2620      2630      2640

tcttaacatt agtgactcca tctagattct gacagttttc atgagaactt gcttttcttt
2650      2660      2670      2680      2690      2700

tctctctcaa gtccttatcc agtattcagc acccttaaca gattagtccc actgctgagt
2710      2720      2730      2740      2750      2760

caggcctctt gcatgaagca gcaatgagaa agacacactt ggccaatggt atcctggagt
2770      2780      2790      2800      2810      2820

aattctcaat gatgccttct ctgtgtttct tcaagacaac tctccttagt gtgagaaaaa
2830      2840      2850      2860      2870      2880

gtccagattt ctcactctct ggccaatgga aggttggtac tcagagcaac ggcaaggagc
2890      2900      2910      2920      2930      2940

```

Figure 2 (suite)

10/53

fragment3.FMO2

```

agagtgcgtg ctttgacgca gttatggttt gcagtggcca ccacattcta cctcatatcc
2950      2960      2970      2980      2990      3000

cactgaagtc atttccaggt gagaccgct gggattccca gctttttgga gtaggtttcc
3010      3020      3030      3040      3050      3060

aggtacttta tatgtagttt ggattgacaa gcaggattca ttgctgcaac tgggcagazc
3070      3080      3090      3100      3110      3120

ttggctcaat aagattgaga cagagctaga aagatgaaag acaccaaaca tcattctttgt
3130      3140      3150      3160      3170      3180

ttctattggc ctctgagtct tcattcacaca tagatctcag agccaacttc cttggaagtc
3190      3200      3210      3220      3230      3240

actaagtccct tggcataatt ttagagaatt cacatcaaac tggttctctg ttggagaggc
3250      3260      3270      3280      3290      3300

ccttttagcc atgtgcctgc gttggccttt ttctaccctg ccaaaccacc agcctttttc
3310      3320      3330      3340      3350      3360

acagggccat actcacacac aaggggagag ctcctagaaa gaaatgcttt gcaagttagt
3370      3380      3390      3400      3410      3420

gatggggaga gaagtgcagg aatagaaccc tgcattccagc tgttcttggt cacccaagtc
3430      3440      3450      3460      3470      3480

tttcttcaga gaacacactt ctttcccaag gcccttagga aaatatgtaa tatagtgggt
3490      3500      3510      3520      3530      3540

catagtccag gcctcatatt agaatcacct ggggagcttc taaagccctg atggcctgga
3550      3560      3570      3580      3590      3600

gacctacccc caaagattca aacactatgg agtagggta gagcaatgaa agtttgctca
3610      3620      3630      3640      3650      3660

ggtgatttta atatacagtc aggattaagg cctgctcatc taaagcaatt gttctcaaat
3670      3680      3690      3700      3710      3720

agagtcacct ggagggtttt tgaaagcaca aattgctagg cccaccctc catatttctg
3730      3740      3750      3760      3770      3780

attcaatagg tgctatggct tgaatgtcct gtccaaaact catattgaga ttaatcccca
3790      3800      3810      3820      3830      3840

atggggcagt atgaagaggt ggggccttta agaggtgatt gagttagtaag agctctgccc
3850      3860      3870      3880      3890      3900

tcaagaatgg attaagccat ttgtggataa atagggtaat ggattattgg gttacacagg
3910      3920      3930      3940      3950      3960

agtggaaactg gtggctttat aagaagagga agagagacct gagctagcat gttagcatgc
3970      3980      3990      4000      4010      4020

ttggctccct caccatacaa tgccctatgc tgccttgga ctcttcagag tccaaaccag
4030      4040      4050      4060      4070      4080

caagaaggct ttcagcagac gcagcccttc aaccttgact tctcagcctc cacaattgtg
4090      4100      4110      4120      4130      4140

tgccagaaga aataacttcc ttccctata aaatattcgg ttccagatat tttgttaaaa
4150      4160      4170      4180      4190      4200

acaatagaag acaaattaa acagtagctc tggcatgagg ctgagaatct gcatttctaa
4210      4220      4230      4240      4250      4260

caccaggcaa tgctgatatt gctggccatg tgaccacact ttgagaacca ataattctaa
4270      4280      4290      4300      4310      4320

gattctttca agcaacccca ccatcaatgg caaatacttt ataaagtcat gtgtttccgt
4330      4340      4350      4360      4370      4380

```

Figure 2(suite)

11/53

fragment3.FMO2

```

gaagtgtaaa agtagtaact aggaaaggac acagaagaag cttgtctgtg attaacccacc
4390      4400      4410      4420      4430      4440

agcaagtcac tgatttacac aatatggaaa ccaactccta tgtgcctggg ttttagtttt
4450      4460      4470      4480      4490      4500

agtttttggg tactttttga aaataagatt gctaaattgt attctaacta ttacacaatt
4510      4520      4530      4540      4550      4560

ataataatag cacttcataa tgtgcttaag aaatatttaa gagtatctga taagtgtatt
4570      4580      4590      4600      4610      4620

tttttttttt tgagatggag tctcactgtc acccagactg gagtgagtg gcacgatctc
4630      4640      4650      4660      4670      4680

agctcactgc aacctccaca acctccatct tccaggttca agcaattctc ctgcctcagc
4690      4700      4710      4720      4730      4740

ctcccaagta gctgggatta caagtgcacg accacccttg gctaattttt gtatttttag
4750      4760      4770      4780      4790      4800

tagagagagc ttcattcatgt tggccaggct ggtttcaaat tccctgacct agttgatccg
4810      4820      4830      4840      4850      4860

cctgccttgg cctcccaaaag tgcctgggatt acagggtgtga gccaccacac ctgcctaat
4870      4880      4890      4900      4910      4920

atgtgatatt aaagggtcaa atgtcattat atagtccaaa atagtatata ataggcaggc
4930      4940      4950      4960      4970      4980

agaagacagt atctgggtcct gctgtgttca tcaccattta tttgtctctg atagagacaa
4990      5000      5010      5020      5030      5040

actgcagccg taagctgcag cctctgaaat aaaaaatcaa cccctttggg cctgtttttt
5050      5060      5070      5080      5090      5100

tgttttgttt ttgtttttgt ttggtgttgt gacagtctca ctctgtcacc cagactggag
5110      5120      5130      5140      5150      5160

tgcaagtgaat caatcagggg tcactgcatt ctttacttcc caagctcaag caatcttccc
5170      5180      5190      5200      5210      5220

acctcagtca cccgagtagc tgggaccaca ggcatgcaca accatgccca gctaattttt
5230      5240      5250      5260      5270      5280

gtattttttg tagatacagg gtttcactat gctgctcagg ctggtctcaa actcctgggc
5290      5300      5310      5320      5330      5340

tcaatcaacc tgcctaggcc tcccaaagcg ctgggattac aggccccacc tggctcggta
5350      5360      5370      5380      5390      5400

cctaaacttt cttatgtgct ttactcctat agagaagagg caaaacaatt attaaactca
5410      5420      5430      5440      5450      5460

gaaaggaaaa gctggcaatg cagttttatt gaaattagct tgacatagtt gctctggagc
5470      5480      5490      5500      5510      5520

tcacagactt ctctcttctt cccctgaag gtatggagag gtttaaaggc caatatttcc
5530      5540      5550      5560      5570      5580

atagccgccca atacaagcat ccagatggat ctgagggaaa acgcatcctg gtgattggaa
5590      5600      5610      5620      5630      5640

tgaggaaactc gggctcagat attgctgttg agctgagtaa gaatgctgct cagggtgtgat
5650      5660      5670      5680      5690      5700

gctctctgct taccatgtac ctggagggga ggaagtgggg atgccatact ggagaacccc
5710      5720      5730      5740      5750      5760

agccatataa tcgcggctcc aatcctcatt aactagttgg ttggtagcgc attgtggcat
5770      5780      5790      5800      5810      5820

catagaaaat ctggaagtca agaaaccact ttacctcta gctctgtcac taaccagcca
5830      5840      5850      5860      5870      5880

```

Figure 2(suite)

12/53

fragment3.FMO2

```

tgaatcctag agtgattcat ttcacttctc tgggagatgg ctccctcatt tttaaaatgg
5890      5900      5910      5920      5930      5940

gaacttttga ccagatgatt ttccatataa gaggcctttc atcaacatgg ctctactgcag
5950      5960      5970      5980      5990      6000

ccttgacctc ctgggctcca atcttctctg catctcagcc tcctgagtag ctgggactac
6010      6020      6030      6040      6050      6060

aggcacatgc cacaccacac tcagctaatt ttcatatatt tgtagagatg agggctcttgc
6070      6080      6090      6100      6110      6120

catgttgccc agggtagtct aaaactcctg aactcaagca atctgcccgc ctccagcctcc
6130      6140      6150      6160      6170      6180

caaagtgcctg ggattacagg catgcacaac cacaccagc caagaggcct tgtttctacc
6190      6200      6210      6220      6230      6240

tggatgttta atgagnnagg tntaatctgt tcatattctg gaggggtggct tttagaaatt
6250      6260      6270      6280      6290      6300

tagtgtgtat ttgaattata tttgaaatat agataacctt cagttaccca aatattatga
6310      6320      6330      6340      6350      6360

aaagaaagat taaatagata gtaggtctct caactaaaat catagatatt taggtgcttc
6370      6380      6390      6400      6410      6420

ctgaggcctt ctaaccactg tcttctttgc acctgctcag gaatgacacc agctgagctg
6430      6440      6450      6460      6470      6480

ccaagagtc aaacattcat tacatgatga tgctgctgac agtggtgggc aggaatagca
6490      6500      6510      6520      6530      6540

aaaactaaac tcctttctgca aggacagacc taggcaaaga agggaaaatc actaaacatc
6550      6560      6570      6580      6590      6600

ctttcccaaa gtattccctc tcaagaaggc ctgaaccaga tgcccaatca ctcttacctt
6610      6620      6630      6640      6650      6660

agctctttca gcctgatgtc tctggccacc cagggcttac catggccctg tgcacaacca
6670      6680      6690      6700      6710      6720

acaaatcatt tccatcctaa gtcttacact ttcaggactc tagatacca gtggcaaaag
6730      6740      6750      6760      6770      6780

ttacaagcaa acatgacacc cgtccagcag gttaatgaag gggttatact gggacctgtc
6790      6800      6810      6820      6830      6840

agagtcattc atcagtcagt tagttagtgc cagcccggga acagagcagg tcactaacac
6850      6860      6870      6880      6890      6900

cggaaagaga cttactagac ccaataagtc ttcactttgt gaaaataaac ctcttgtcac
6910      6920      6930      6940      6950      6960

ttatcacctc agtgtgaaga acaagtgagg aggcaggaaac tgtgacagcc tggagaagag
6970      6980      6990      7000      7010      7020

cagagctgga aaatgagagt accagctcta ggctctttca tgctacgaat acccgcaaag
7030      7040      7050      7060      7070      7080

ccttaggaac agagtgtaat ggggcagtat gtgaggagct aatatagcag tcagccaagt
7090      7100      7110      7120      7130      7140

gaagatccat cctagactac ttcacgttgt cagaccagtg atttgattt agatctcttc
7150      7160      7170      7180      7190      7200

attccaaaga tatcaaatct tagatggcaa gaaccagttc cttgtatggg tcttgcccta
7210      7220      7230      7240      7250      7260

caggaagact tatgggtgtg gattcaatat taagaaacta ccttggtctt atttgcatgc
7270      7280      7290      7300      7310      7320

cttacagctt cttaaacaat cttttgcaca gagtgcacaa gactttgttt ccatctccct
7330      7340      7350      7360      7370      7380

```

Figure 2(suite)

13/53

fragment3.FMO2

```

ctatcagtgt aaatgccact agatgcccc tttttaggag gtacttcact ttgagggtcaa
7390      7400      7410      7420      7430      7440

tcattctttta aacagagcct cagtaaatte tggggctatg catgtgatac atcacctaca
7450      7460      7470      7480      7490      7500

taatagattc ctcctaaata taatgtttata atcatacatt tccaggatta tactcattca
7510      7520      7530      7540      7550      7560

cctgcactaa tctcttcaat atttattaga gtaacaacat aaatctataa ctatgataaa
7570      7580      7590      7600      7610      7620

acctctttaca cagagtaata tactctcaag ccttctgtga aaagactaac cagagacttt
7630      7640      7650      7660      7670      7680

acaggagcta tacatgctag gaacggaact aggcgcattt gcaaaacttg aaattacaac
7690      7700      7710      7720      7730      7740

ctgaactcac caaaattctg agtgtgcact gctctgttaa aagaaattca ccttcataag
7750      7760      7770      7780      7790      7800

gttacagcac cctctaccac aatccaaaag caccactcaa gatcatatgg gatggtgctg
7810      7820      7830      7840      7850      7860

catcattgta ttagtccatt ctcaacgctg ctatgtagac atacccgaga ctgggtaatt
7870      7880      7890      7900      7910      7920

cataaagaaa agagggttaa ttgacttaca gtttggcatg gctggggaag cctcaggaaa
7930      7940      7950      7960      7970      7980

ctaacaatca tgacggaatg cacctcttta cgaggctgca ggagaaagaa tgagagcgac
7990      8000      8010      8020      8030      8040

tggggaaccc cttataaaac catcagatct cgtgagaact tactccctat taggagaaca
8050      8060      8070      8080      8090      8100

gcatggcaga aacctcccc atgattgaat tatctccacc tggctcctgc cttgacacgt
8110      8120      8130      8140      8150      8160

ggggattatt ataatttaag gtgagatgtg ggtagggaca cacagccaaa ccatattagt
8170      8180      8190      8200      8210      8220

catttacata cttctgacca aaaaccaa atctggcctt tgacctaaaa catgcgtctc
8230      8240      8250      8260      8270      8280

agagaaagca gcctgagcct aaatcctcat gtttctctca ctgttgacgc tagtgctatt
8290      8300      8310      8320      8330      8340

aaggcaggtt agaccacct gctgtaggga gggtcacaac agaaaaagag tgaatcaaac
8350      8360      8370      8380      8390      8400

gggcagagca taccatttga aacatgggtt gctcctgaga aagaagagg gacagtaagt
8410      8420      8430      8440      8450      8460

aatggaaaga gacactaatg aaaatatttt tgtatcta atctaataa agtattgcc
8470      8480      8490      8500      8510      8520

agtcagccta taagggaac ggcaggagaa attcagaaca taggtatata ccacacacag
8530      8540      8550      8560      8570      8580

accagcaata taggaatgct tggatatagg gctacttcac aagctaggaa tgtaaggccc
8590      8600      8610      8620      8630      8640

atccccacaa aatttgtctc caaattctgg tttactccag acataaggca ctgtatgaaa
8650      8660      8670      8680      8690      8700

ctcctctctt ccagcctaac tttataactt aacagctagc agtacttatc acttgccagg
8710      8720      8730      8740      8750      8760

caatatttca agtactttat atataccacc tcattttaac tacacaagaa tgccatgagg
8770      8780      8790      8800      8810      8820

```

Figure 2 (suite)

fragment3.FMO2

```

taggtactgt taataccccc attttacaga gagagaaact gaggcacaga gagattgaaa
8830      8840      8850      8860      8870      8880

taattcaacc atggcaacac agattgaaat agttcaccca cagtagtggt attgggattc
8890      8900      8910      8920      8930      8940

aaacccaagc agtctgtatc caaacctctc aagtaaattg gttaccttgc aagtgaatct
8950      8960      8970      8980      8990      9000

tatgtgttta tcaagtatag ccttaaacaa aaacttattg catgggtatgt aaaaatttaa
9010      9020      9030      9040      9050      9060

gaagcagttc aagtatgcat ttggccaatg ggggagtaac agcaaacaca gcaaaatata
9070      9080      9090      9100      9110      9120

catttgaaaa gagattaaat gtacattttg gaaacaaggg aaatcttaat aaacaaggta
9130      9140      9150      9160      9170      9180

aagaatacac ctgaaagagg attcagatgt gcacttgaag agaaagagaa tcacagtata
9190      9200      9210      9220      9230      9240

agttcagagt ttttaacttt taaaatacat tacaagcact gtgtctcatg cctgtaatcc
9250      9260      9270      9280      9290      9300

cagcactttg ggaggctgcg gcaggaggat tgcttaagcc cagaaatttg agaccgacct
9310      9320      9330      9340      9350      9360

gggcaacata atgagaccgt ctctacaaaa aaattgtttg aattagctgg atgtggtggt
9370      9380      9390      9400      9410      9420

acatgtctga tactgagggt ggaggatcac ttgagcctgg gaggtcgaga ctgcaatgag
9430      9440      9450      9460      9470      9480

ctatgactgc acaactgcag tccagcctga gtgacagagc aagaccctgt ctcacacaca
9490      9500      9510      9520      9530      9540

cacacacaca cacacacaca cacacacaca aaataaagtc ttttaagtat ggaagggaaga
9550      9560      9570      9580      9590      9600

ttatttcccc tggtattctc catccaggga tattcagatg catatacact tatactttgtg
9610      9620      9630      9640      9650      9660

tagtcactag gctataatcg cacatttcca aggattataa tcattctacc tgcactatag
9670      9680      9690      9700      9710      9720

aagaaactta ggtgagtggg aaacatgaga ggaggaggag aggaactttc tcttaaggag
9730      9740      9750      9760      9770      9780

cagcaaacca caactgtaaa catgggaaag acttggtgat tttatcatca gaggtagccc
9790      9800      9810      9820      9830      9840

aaagactttc tcgtgtctcc atgaagttct caagattttg ttgcagtctt cctgcatcag
9850      9860      9870      9880      9890      9900

tgtaaatgcc actgggtacc cctatttagg aggtacttta cattgaggtc aatcatcttt
9910      9920      9930      9940      9950      9960

aaaacagaac ctctgtaaat tctggggcta cacatgtgat acatgacctt catagtagat
9970      9980      9990      10000      10010      10020

tcctcctaaa cgggacaatg ccctaattta aactgcattt ctttttgctt gccagggttt
10030      10040      10050      10060      10070      10080

tatcagcacc aggcattggc cctgggtcat gagccgtatc tctgaagatg gctatccttg
10090      10100      10110      10120      10130      10140

ggactcagtg ttccacaccc ggtttcgttc tatgctccgc aatgtactgc cacgaacagc
10150      10160      10170      10180      10190      10200

tgtaaaatgg atgatagaac aacagatgaa tcggtggttc aaccatgaaa attatggcct
10210      10220      10230      10240      10250      10260

```

Figure 2(suite)

15/53

fragment3.FMO2

```

tgagcctcaa aacaagtaga gttattttgc ttttttaatg gtataactcg tgggtgagcaa
10270      10280      10290      10300      10310      10320

agttgtctga aggtgtctcc cttaacaaag attcaaattg ctaacacggt agttaaaact
10330      10340      10350      10360      10370      10380

acaatctaac aatatgagta tcttataggt cctggagttt agctttctaaa tttggactga
10390      10400      10410      10420      10430      10440

acccctttta aaaatacgtg agaagatgaa gcagaagtgt tataagctgc tccagaaaagc
10450      10460      10470      10480      10490      10500

aaaactaggg gagaactttc taatacccag agttatctaa cattggagaa aactgttttca
10510      10520      10530      10540      10550      10560

agagattacg acctgccttt cagaggggtg tgggtgggaaa catgtaattc tccatctaat
10570      10580      10590      10600      10610      10620

aatttatgct ttgctaaccg tatagcatga aggttcttcc catgggaaac ctttgaaaac
10630      10640      10650      10660      10670      10680

acattccttt ttcttttgcta aaagacaaat ctctgttgaa cgtcaaagtt atatgtcagt
10690      10700      10710      10720      10730      10740

gattttaagca caagcaaatg ttatgaatgg ttcttttgct ttagttgtta caggcttctt
10750      10760      10770      10780      10790      10800

cccttaaaaa aacagaagag ctttagaatc ttttaacaaa tgcctgccgt gcaactacca
10810      10820      10830      10840      10850      10860

tatttctaaga tctgacataa gtgccacgta tcgtctatta aaaaaagaaa aagaaaaatgt
10870      10880      10890      10900      10910      10920

tctcaaatct acaaaaaaaaa taagcggact ttgcatcaac atccatgcta ttactaacag
10930      10940      10950      10960      10970      10980

agactccatg gatatttggg attaacaaat atcaccaaac ctaattttat acattaattt
10990      11000      11010      11020      11030      11040

tcacattgat cccttcatag atttcaaaac tagtggaat ttagcaaat ttttcttatg
11050      11060      11070      11080      11090      11100

atcaaataag ggttaaataa aacagcaaaa taataaaagc tagatagcat gaaaaagggt
11110      11120      11130      11140      11150      11160

aaaaacagaa atggtataat aaccaccata atacttgggg attgaccata ggcacaggca
11170      11180      11190      11200      11210      11220

ttttgtctaa gcccttgggg atgcttctct ccttaaaatc tctttcactc acgttgccta
11230      11240      11250      11260      11270      11280

catgttttcc cttatttatt gacaagagat atttgtgaca tgagaattaa gtcagaaaat
11290      11300      11310      11320      11330      11340

aaggatttgc acagacaacc agttaagtta gagttttaca gatatttgaa aagccctttt
11350      11360      11370      11380      11390      11400

attttcagag ccgtacccca aaaaatatcaa gaggggtcaa gattcctcag caaatgatcc
11410      11420      11430      11440      11450      11460

ttcagaatgt ttttcttctg tatgtctcag atacattatg aaggaacctg tactaaatga
11470      11480      11490      11500      11510      11520

tgatgtccca agtcgtctac tctgtggagc catcaagggt aaatctacag tgaaagagct
11530      11540      11550      11560      11570      11580

cacagaaact tctgccatct ttgaggatgg aacagtggag gagaacattg atgtcatcat
11590      11600      11610      11620      11630      11640

ttttgcaaca ggatatagtt tctcttttcc cttccttgaa gattcactcg ttaaagtaga
11650      11660      11670      11680      11690      11700

gaataatatg gtctcactgt ataaatacat attccccgct cacctggaca agtcaacctt
11710      11720      11730      11740      11750      11760

```

Figure 2(suite)

16/53

fragment3.FMO2

```

cgctgcatc 11770 ggtctcatcc 11780 agcccctagg 11790 ttccattttc 11800 ccaactgctg 11810 aacttcaagc 11820
tcgttgggtg 11830 acaagagttt 11840 tcaaaggtaa 11850 gtgtgtaggc 11860 aggtgagtg 11870 ctaagcggtt 11880
cagatctggt 11890 gaagtttatc 11900 aataatgata 11910 agaagggtgc 11920 ctgagataaa 11930 aaggttgcca 11940
agaaaaagt 11950 tgacaacctt 11960 ggctgctctc 11970 acaagactaa 11980 cattctaaaa 11990 agttactgga 12000
gaattcaaa 12010 aataacaaat 12020 acaggaat 12030 agtaataata 12040 aatacctgca 12050 atcatccttc 12060
taaaatatta 12070 gacagtcaag 12080 agaatttcaa 12090 ctggcataaa 12100 gctaagtgca 12110 tgttaacttt 12120
tctttgaatc 12130 gtgagagata 12140 agtttaagaa 12150 aaagatctgt 12160 ctccctgggtt 12170 tacctctgtg 12180
ttgtttaaaa 12190 attcctcagc 12200 atatctgcaa 12210 atcaatttaa 12220 ctcttaatac 12230 ttgagcagct 12240
caacctcaca 12250 aatccctaca 12260 agttataaaa 12270 ttattaaaag 12280 gtttctttct 12290 ggtgtgtgt 12300
gtagcacttc 12310 atactcctca 12320 gaacggtgtt 12330 acctccctgc 12340 ctccagggtt 12350 caattctgtt 12360
cagcaaaagc 12370 ttactgaata 12380 ccttgccctg 12390 tgctgggaac 12400 tgggtgggaca 12410 gagagaaatt 12420
taaacagatc 12430 atttcaacat 12440 aacatgacaa 12450 atgctttgat 12460 tgaataatat 12470 atggagtgtt 12480
cagggaagga 12490 gagaaagggc 12500 acttatcatg 12510 gtagaataag 12520 ggaagggcac 12530 atgataaagg 12540
aaaaagtcct 12550 ggataactgc 12560 atttctcagg 12570 ggcagaaaag 12580 gggattgcct 12590 gaacaaaagc 12600
atagagtcaa 12610 tgatgcata 12620 ggaagggcac 12630 atgctatttg 12640 acattgctag 12650 agcatgacgt 12660
atgaggcaga 12670 gagagatgag 12680 ccattactct 12690 tggagaagaa 12700 ggagacagga 12710 cacaggaatt 12720
ttttaagaca 12730 tgctatggag 12740 cttagattat 12750 aaattataga 12760 tcagttcttc 12770 ccaaatatgg 12780
ctacatatga 12790 aaatcatctg 12800 atggatcctt 12810 agggaccctg 12820 attaagtaag 12830 actggccaag 12840
ggacctggaa 12850 tctgcatctt 12860 agaaagctct 12870 tcagcccggt 12880 gcaccaatga 12890 aggtttataa 12900
gcaaggaaac 12910 ggcattagca 12920 gatttacact 12930 tcagatagat 12940 tgtttcagca 12950 gtagtgtgga 12960
atatagatct 12970 gaaagtgggg 12980 aaagactaca 12990 gcctcaggga 13000 tgaaagagaa 13010 agctactgaa 13020
atagcctatg 13030 ctaaaatatg 13040 atgcatcctg 13050 ggcaggga 13060 gagatacaaa 13070 gtggaaagga 13080
agccataaat 13090 gtgagaaatc 13100 attaaggga 13110 aaatcagcat 13120 gacattatga 13130 ttggttcaat 13140
gtgggaaagt 13150 cagagaaata 13160 gagaggaatc 13170 taggaggact 13180 tacagatctc 13190 tggcattgga 13200

```

Figure 2(suite)

17/53

fragment3.FMO2

```

aaccagggtgg acagtagtgc tgtgaatata gaggggggtgt gcagaaaatg atgcaagtctt
13210      13220      13230      13240      13250      13260

ggacaggagg gcttcagtga ggagctcagg tctggactac ttgaacatga gatgtctgat
13270      13280      13290      13300      13310      13320

gactctaggc aaggggactt gaccatattt caacacatcc aaagctcagg ggacacttgt
13330      13340      13350      13360      13370      13380

gggcaggcga tggagtcagt agcacacagt aataacttct gcacaaatct tccctatctt
13390      13400      13410      13420      13430      13440

ctactgccct actctcatct ctcaccaggt ttatttcaac agcctcttta ctggtctccc
13450      13460      13470      13480      13490      13500

cagcttttggg cttgcctccc tggagtcagt tttcctaagt tcagcagcca gacagatctt
13510      13520      13530      13540      13550      13560

tccaaaaaat aaatctgagt ttctcacttc attcagaata ctcttccact gatttgattt
13570      13580      13590      13600      13610      13620

ggggcctcct gtcaccttca ggatagagcc caaaccacta gtcattggtg ccaggctccc
13630      13640      13650      13660      13670      13680

agacacactt cctttttcca gcctcttctc ttggcctctc ccacttgtag tccatgccgt
13690      13700      13710      13720      13730      13740

agactgtgca ccctggacag tgtcacatag agtgctatgg ggggtggcacc ccctgaagtt
13750      13760      13770      13780      13790      13800

caacagcacg gaagccctga ctggtatgac atggttcaat gtccagagtt taattttaag
13810      13820      13830      13840      13850      13860

aatcaacaac tagacaaagt aatgatattg actcaaacct actattcaaa ccaacctttt
13870      13880      13890      13900      13910      13920

attccttagg cttgtgtagc ctgccctcag agagaactat gatgatggac attatcaaaa
13930      13940      13950      13960      13970      13980

ggaatgaaaa aagaattgac ctgtaagaat tttttttaat tctttacatg aagcagtgtt
13990      14000      14010      14020      14030      14040

tctcaaagta cagtgatcta actacttaca agaaccacct agctgcctga taaaaatgcaa
14050      14060      14070      14080      14090      14100

atctctgggc tatagcccag atgattgaat cagaaactcc gtgtgtgagg ctaaaaagtt
14110      14120      14130      14140      14150      14160

gcatttttat cttcttccta agcgattctt atacatacta ggttaagaac caaatactta
14170      14180      14190      14200      14210      14220

aagataagaa ttgtacccaa tcagagcact tctccttggc ttaatttcat ttcagtttta
14230      14240      14250      14260      14270      14280

tatgatgcct atgtcagatt ccataacttc tcaagccacc tacactctgt ggttagagag
14290      14300      14310      14320      14330      14340

ggaatgggat gagacagtgg tgggtgatgt agcttgaata gctgtgaaaa gttagagaat
14350      14360      14370      14380      14390      14400

ccccatcaga ataaattagg aaggggttgg tgtgaagggt caaggatttg tactttgtga
14410      14420      14430      14440      14450      14460

tgaggtaaaa tgaggttcaa cagtgatcga gtacccttgg aaagttgatt tggggcttac
14470      14480      14490      14500      14510      14520

atcagggtgta aagagttttc tcatgttcaa attcaaatat acctaagatt gattgagtat
14530      14540      14550      14560      14570      14580

ctactatacg ccatccagac tgccaggtag tttagtaatt taacaagcaa atattaagca
14590      14600      14610      14620      14630      14640

```

Figure 2 (suite)

18/53

fragment3.FMO2

```

tctcctttga gcaagacacc aagctatgct ttcatatgca ttatctcatg aattcctgca
14650      14660      14670      14680      14690      14700

gccgccttgg ctagcatgta cttgcctgga gatttgccac cgcttaaaaa atgccaaaca
14710      14720      14730      14740      14750      14760

atggttacca atcttgtcac atttctagag catccatgaa ttcattggctc tttatttgag
14770      14780      14790      14800      14810      14820

ggcgtattct caatctgaga tatgagcctc ctggtatgat aaactcaaac tttccaccag
14830      14840      14850      14860      14870      14880

agattcattg aaaactcatt cacatattca ctcatctcctt cattccttta gcagttttga
14890      14900      14910      14920      14930      14940

atgcctaata ttctagaaaa cttagaacat tctgtgaaca ttccctttttt actttcttca
14950      14960      14970      14980      14990      15000

ctaaggtttg gagaaagcca gagccagacg ttgcagacca attatgttga ctacttggac
15010      15020      15030      15040      15050      15060

gagctcgccct tagagatagg tgcgaagcca gatttctgct ctctcttgtt caaagatcct
15070      15080      15090      15100      15110      15120

aaactggctg tgagactcta tttcggaccc tgcaactcct attagtatcg cctgggtggg
15130      15140      15150      15160      15170      15180

cctgggcaat gggaaggatt cagaaatgcc atcttcaccc agaagcaaa aatactgaag
15190      15200      15210      15220      15230      15240

ccactcaaga ctcgggccct gaaggattca tctaatttct cagtttcttt tctgttgaaa
15250      15260      15270      15280      15290      15300

atcctggggc ttcttctgtt tgttgtggcc tttttttgcc aacttcaatg gtcctagtca
15310      15320      15330      15340      15350      15360

gcataatgct ttgggcttta ttatctgttc agtcactacc tcctaaagaa aaaaaaaaaag
15370      15380      15390      15400      15410      15420

gctagaagaa aaaacattac attcatgttc taattataga ttttagagtt aggtagtaca
15430      15440      15450      15460      15470      15480

ggtaaggggg aaattgtaaa gaattagcag aattaggcat atgtacaaaa ccaaaatttt
15490      15500      15510      15520      15530      15540

gtcatgaaat tttgcctttc cagcttctcc tcagttcacc aaagttacca aaatgtaaaa
15550      15560      15570      15580      15590      15600

taaaataaga ctggctcagg taagtagtgc tgccaaccct gatatagggg agttgtatgg
15610      15620      15630      15640      15650      15660

aaaaatagta gaattacaca gcatgaaaag cagcccatgg tttaaattat tggacaattt
15670      15680      15690      15700      15710      15720

aaattgtggg taaatattta aaactcctga acaatgtttc tgatgggtctt ctatccaccc
15730      15740      15750      15760      15770      15780

tacttggtaa caaagtcttc agatgttagg tcatgtttca tttgctcagt cggggatcac
15790      15800      15810      15820      15830      15840

tcaaaactac tagacaaaaa agtgagagga tagattttag aaacatcagt gatgtcaga
15850      15860      15870      15880      15890      15900

taaactttta ggacctcata ttaagagcta agcaaatggc cacatttcctt atattttgac
15910      15920      15930      15940      15950      15960

agagatactg ctggaaaaat taaaattaaa atgccataat agctaccta caaatatata
15970      15980      15990      16000      16010      16020

tgtttaatgt ttatcatagg ccagacattg tgctatgtgc atatcatatg tattatttca
16030      16040      16050      16060      16070      16080

tttaattctc acaacaattc tgtgaaatgg taacagctat tatagtcatt tcacagatga
16090      16100      16110      16120      16130      16140

```

Figure 2 (suite)

19/53

fragment3.FMO2

```

tgaaactaag attcagagca gctgatcttg tgaggcagct ggaattggaa ctccagatttg
16150      16160      16170      16180      16190      16200

ttgaactcta gaactaaaga tcataatggt gtcttgtaat atattttattt acaaaacact
16210      16220      16230      16240      16250      16260

tcattattta taaagaattt actaacagtt tatctttattt ataccatac atctgctact
16270      16280      16290      16300      16310      16320

ttggggaggcc ctttacatag aaaacagcat tctttttgct aaatatgacc aaattacttt
16330      16340      16350      16360      16370      16380

tattttataat ttttgattta tttttcagct agatctaaaa agcatctgaa ggaattttaca
16390      16400      16410      16420      16430      16440

atgaaagata cctatgcaat aacatttagg ataactcttg acatttttggg aaaataagaa
16450      16460      16470      16480      16490      16500

ttgaggaaaa aaagtgtatc tttcaagtag atgcaaagca ttataatgat tgacacttgt
16510      16520      16530      16540      16550      16560

atctaactcc agtcttacag ataactaagg caaaaagcta aataaacaat atgtaacctc
16570      16580      16590      16600      16610      16620

taacatttgg taaaagggaag tatactggtc tgtttagcaga gacaaaacttt ttttagaatt
16630      16640      16650      16660      16670      16680

gaagtctgaa acaaaacaaa gcaattcaat gtcaatagac attaagcaac ataatagaca
16690      16700      16710      16720      16730      16740

aacatctcct aagggaacat ttgttacagc tgctccttcc ctgaactgtg ctttggaga
16750      16760      16770      16780      16790      16800

taagctctgt cctgagtgca aaccaagccc ttccaagaga gaacaaagggt cagagatgtt
16810      16820      16830      16840      16850      16860

gaagattcca gcaaatttct cctctttattt ctaccaagcc tttgtgaaca ttgctcttca
16870      16880      16890      16900      16910      16920

ttttggcctg tacttctccc tcagggacgt agaacaatgg aatgtcagtc agtctctgta
16930      16940      16950      16960      16970      16980

gttaaaactt tttctttaaa attcaattaa gttacttctc cctcaggac gtagaacaat
16990      17000      17010      17020      17030      17040

ggaatgtcag tcagtctctg tagttaaaac ttttctttta aaattcaatt aagttacacc
17050      17060      17070      17080      17090      17100

agaatttaca ggcaagattt tttttttcat tgctcccata agcaaatttg ttttaaaata
17110      17120      17130      17140      17150      17160

attgtaaatg aggtatatac ttagtctctg gttaaaaaat atattgcttt gtttaagtatt
17170      17180      17190      17200      17210      17220

aaagattatt tgtaagtcac tgtattaata atactaataa aatttatcaa gcctttatag
17230      17240      17250      17260      17270      17280

caagggtcag tgaattacca ctgcctgttg gccaaatcta gctcactatc tgtttttgta
17290      17300      17310      17320      17330      17340

aataaaattt tataatagta cacagccaca ctcatctatt tattttctgt ggttgctttc
17350      17360      17370      17380      17390      17400

aagctacaat tgtagagttg ggtagtcgca acagaatctc tgtggccccc aaggctaaaa
17410      17420      17430      17440      17450      17460

tattttacatt ctccaccatt acagaaaaag tttgataatt cctgctttat aatatgtaag
17470      17480      17490      17500      17510      17520

gcattgtccc attttgcata acctgcctta tttcatcatt atcactaccc atttagtagc
17530      17540      17550      17560      17570      17580

tatggttgtt atcttacttc tacagtggaa gagattgaaa agcattttgtc aggttaatgc
17590      17600      17610      17620      17630      17640

taaattcagtg cggaaatata gctc
17650      17660

```

Figure 2(suite)

20/53

FIGURE 3A

FMO2	position sur le gène	position sur le mRNA
Exon 1 :	2001 - 2056	1 - 56
Exon 2 :	2405 - 2542	57 - 194
Exon 3 :	10026 - 10214	195 - 383
Exon 4 :	13341 - 13503	384 - 546
Exon 5 :	16036 - 16178	547 - 689
Exon 6 :	20558 - 20757	690 - 889
Exon 7 :	21972 - 22327	890 - 1245
Exon 8 :	24411 - 24483	1246 - 1318
Exon 9 :	25487 - 25899	1319 - 1731
ATG :	2411 - 2413	63-65
Stop:	25836-25838	1668 - 1670

FIGURE 3B

FMOx	position sur le gène	position sur le mRNA
Exon 1 :	2001 - 2138	1 - 138
Exon 2 :	6961 - 7149	139 - 327
Exon 3 :	10144 - 10306	328 - 490
Exon 4 :	11413 - 11555	491 - 633
Exon 5 :	13347 - 13546	634 - 833
Exon 6 :	15697 - 16052	834 - 1189
Exon 7 :	17930 - 18002	1190 - 1262
Exon 8 :	24838 - 25180	1263 - 1605
CDS :	2006-25180	6 - 1605

homologie FMO2

	Longueur	% homologie Macaque/homme Acides Amines	% homologie Macaque/homme ADN
exon 1 (5' UTR)	64	-	95.3
exon 2	137	100	96.5
exon 3	188	98	96.8
exon 4	162	96.7	96.9
exon 5	142	95.8	96.5
exon 6	199	95.4	97
exon 7	355	98.3	97.7
exon 8	72	96	97.2
exon 9 (3' UTR)	413	93	95
Pourcentage Global		96	96.7

Figure 4

22/53

Tableau FMO2

Variation homme/macaque sur FMO2

position mRNA macaque	base macaque	base humain	amino acid
56	A	G	non codant
71	A	G	-
83	C	T	-
104	G	A	-
197	G	T	Lys -> Asn
218	C	T	-
266	T	C	-
284	C	T	-
344	C	T	-
360	T	C	-
404	G	A	-
455	T	C	-
482	T	C	-
499	C	G	Ser -> Thr
510	T	A	Ile -> Phe
548	C	G	Ile -> Met
604	T	C	Ser -> Phe
629	C	T	-
650	C	A	-
676	G	A	Asn -> Ser
725	T	C	-
729	G	A	Val -> Ile
743	T	C	-
758	G	A	-
811	T	C	Val -> Ala
844	A	G	Arg -> Gln
995	T	C	-
1085	T	C	-
1121	G	A	-
1133	A	C	-
1145	G	C	Glu -> Asp
1155	T	C	Phe -> Leu
1157	T	C	-
1160	A	G	-
1251	C	A	Ser -> His
1252	A	G	Ser -> His
1370	T	C	-
1448	G	C	-
1450	T	A	Tyr -> Phe
1473	C	N	-
1484	A	G	-
1486	C	T	-
1509	G	N	-
1510	C	N	-
1514	G	A	-
1516	G	A	-
1535	A	G	-
1541	G	A	-
1556	A	C	-
1567	T	C	-
1590	C	T	-
1598	C	T	-
1623	G	C	-
1646	C	T	-
1677	T	C	-
1678	G	A	-

Figure 5

	10	20	30	40	50	60
mRNA _{Pot}	AACCAAGGGGAGAAA	ACTATTCTGTCAAAGAGACGGTGCCAA	AGGCAAAAACA	AGGAGG		
	:	:	:	:	:	:
Macaque	GGCACGAGGAAA	ACTATTCTGTGTC-AAGAGACGGTGCCAA	AGGCAAAAACA	AGGAAC		
	10	20	30	40	50	
	70	80	90	100	110	120
mRNA _{Pot}	TGATGGCAAAGAAGGTAGCTGTGATTGGAGCTGGGGTCAGTGGCCTAATT	TCTCTGAAGT				
	:	:	:	:	:	:
Macaque	TGATGGCAAAGAAGGTAGCTGTGATCGGAGCTGGGGTCAGTGGCCTGATT	TCTCTGAAGT				
	60	70	80	90	100	110
	130	140	150	160	170	180
mRNA _{Pot}	GCTGTGTGGATGAGGGACTTGAGCCCACTTGCTTTGAGAGA	ACTGAAGATATTGGAGGAG				
	:	:	:	:	:	:
Macaque	GCTGTGTGGATGAGGGACTTGAGCCCACTTGCTTTGAGAGA	ACTGAAGATATTGGAGGAG				
	120	130	140	150	160	170
	190	200	210	220	230	240
mRNA _{Pot}	TGTGGAGGTTCAAAGAGAATGTGGAAGATGGCCGAGCAAGTATCTATCAATCTGTCGTTA					
	:	:	:	:	:	:
Macaque	TGTGGAGGTTCAAAGAGAAGGTGGAAGATGGCCGAGCAAGCATCTATCAATCTGTCGTTA					
	180	190	200	210	220	230
	250	260	270	280	290	300
mRNA _{Pot}	CCAACACCAGCAAAGAAATGTCCTGTTTTAGTGACTTTCCAATGCCTGAAGATTTTCCAA					
	:	:	:	:	:	:
Macaque	CCAACACCAGCAAAGAAATGTCCTGTTTTAGTGACTTTCCAATGCCGAAGATTTTCCAA					
	240	250	260	270	280	290
	310	320	330	340	350	360
mRNA _{Pot}	ACTTCCTGCATAATTCTAAACTTCTGGAATATTTTCAGGATTTTTTGCTAAAAAATTTGATC					
	:	:	:	:	:	:
Macaque	ACTTCCTGCATAATTCTAAACTTCTGGAATATTTTCAGGATTTTTTGCCAAAAAATTTGATC					
	300	310	320	330	340	350
	370	380	390	400	410	420
mRNA _{Pot}	TGCTAAAATATATTTCAGTTCCAGACAACCTGTCTTAGTGTGAGAAAATGTCCAGATTTCT					
	:	:	:	:	:	:
Macaque	TGTTAAAATATATTTCAGTTCCAGACAACCTGTCTTAGTGTGAGAAAGTGTCCAGATTTCT					
	360	370	380	390	400	410
	430	440	450	460	470	480
mRNA _{Pot}	CATCCTCTGGCCAATGGAAGGTTGTCACTCAGAGCAACGGCAAGGAGCAGAGTGCTGTCT					
	:	:	:	:	:	:
Macaque	CATCCTCTGGCCAATGGAAGGTTGTCACTCAGAGCAATGGCAAGGAGCAGAGTGCTGTCT					
	420	430	440	450	460	470
	490	500	510	520	530	540
mRNA _{Pot}	TTGACGCAGTTATGGTTTGCAGTGGCCACCACATTCTACCTCATATCCCAGTGAAGTCAT					
	:	:	:	:	:	:
Macaque	TTGATGCAGTTATGGTTTGCAGTGGCCACCACATTTCTACCTCATATCCCAGTGAAGTCAT					
	480	490	500	510	520	530

FEUILLE DE REMPLACEMENT (REGLE 26)

Alignement-mRNAFMO2

```

      550      560      570      580      590      600
mRNApot TTCCAGGTATGGAGAGGTTCAAAGGCCAATATTTCCATAGCCGCCAATACAAGCATCCAG
          : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Macaque TTCCAGGTATCGAGAGGTTCAAAGGCCAATATTTCCATAGCCGCCAATACAAGCATCCAG
      540      550      560      570      580      590

      610      620      630      640      650      660
mRNApot ATGGATCTGAGGGAAAACGCATCCTGGTGATTGGAATGGGAAACTCGGGCTCAGATATTG
          : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Macaque ATGGATTGAGGGAAAACGCATCCTGGTGATCGGAATGGGAAACTCGGGCTCCGATATTG
      600      610      620      630      640      650

      670      680      690      700      710      720
mRNApot CTGTTGAGCTGAGTAAGAATGCTGCTCAGGTTTTTATCAGCACCAGGCATGGCACCTGGG
          : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Macaque CTGTTGAGCTGAGTAAGAGTGCTCAGGTTTTTATCAGCACCAGGCATGGCACCTGGG
      660      670      680      690      700      710

      730      740      750      760      770      780
mRNApot TCATGAGCCGTATCTCTGAAGATGGCTATCCTTGGGACTCAGTGTTCCACACCCGGTTTC
          : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Macaque TCATGAGTCGTGTCTCTGAAGATGGTTATCCTTGGGACTCGGTGTTCCACACCCGGTTTC
      720      730      740      750      760      770

      790      800      810      820      830      840
mRNApot GTTCTATGCTCCGCAATGTACTGCCACGAACAGCTGTAAAATGGATGATAGAACAACAGA
          : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Macaque GTTCTATGCTCCGCAATGTACTGCCACGAACAGTTGTAAAATGGATGATAGAACAACAGA
      780      790      800      810      820      830

      850      860      870      880      890      900
mRNApot TGAATCGGTGGTTCAACCATGAAAATTATGGCCTTGAGCCTCAAAACAAATACATTATGA
          : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Macaque TGAATCAGTGGTTCAACCATGAAAATTATGGCCTTGAGCCTCAAAACAAATACATTATGA
      840      850      860      870      880      890

      910      920      930      940      950      960
mRNApot AGGAACCTGTACTAAATGATGATGTCCAAGTCGTCTACTCTGTGGAGCCATCAAGGTGA
          : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Macaque AGGAACCTGTACTAAATGATGATGTCCAAGTCGTCTACTCTGTGGAGCCATCAAGGTGA
      900      910      920      930      940      950

      970      980      990      1000      1010      1020
mRNApot AATCTACAGTGAAAGAGCTCACAGAACTTCTGCCATCTTTGAGGATGGAACAGTGGAGG
          : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Macaque AATCTACAGTGAAAGAGCTCACAGAACTTCTGCCATTTTTGAGGATGGAACAGTGGAGG
      960      970      980      990      1000      1010

      1030      1040      1050      1060      1070      1080
mRNApot AGAACATTGATGTCATCATTTTTGCAACAGGATATAGTTTCTCTTTCCCTTCCTTGAAG
          : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Macaque AGAACATTGATGTCATCATTTTTGCAACAGGATATAGTTTCTCTTTCCCTTCCTTGAAG
      1020      1030      1040      1050      1060      1070

      1090      1100      1110      1120      1130      1140
mRNApot ATTCACCTCGTTAAAGTAGAGAATAATATGGTCTCACTGTATAAATACATATTCCCCGCTC
          : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Macaque ATTCACCTGTAAAGTAGAGAATAATATGGTCTCACTGTATAAGTACATATTCCCAGCTC
      1080      1090      1100      1110      1120      1130

```

Figure 6(suite)

25/53

Alignement-mRNAFMO2

```

      1150      1160      1170      1180      1190      1200
mRNApot ACCTGGACAAGTCAACCCCTCGCGTGCATTGGTCTCATCCAGCCCCTAGGTCCATTTTCC
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Macaque ACCTGGAGAAGTCAACCTTTGCATGCATTGGTCTCATCCAGCCCCTAGGTCCATTTTCC
      1140      1150      1160      1170      1180      1190

      1210      1220      1230      1240      1250      1260
mRNApot CAACTGCTGAACCTTCAAGCTCGTTGGGTGACAAGAGTTTTCAAAGGCTTGTGTAGCCTGC
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Macaque CAACTGCTGAACCTTCAAGCTCGTTGGGTGACAAGAGTTTTCAAAGGCTTGTGTACCTGC
      1200      1210      1220      1230      1240      1250

      1270      1280      1290      1300      1310      1320
mRNApot CCTCAGAGAGAACTATGATGATGGACATTATCAAAAGGAATGAAAAAGAATTGACCTGT
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Macaque CCTCAGAGAGAACTATGATGATGGACATTATCAAAAGGAATGAAAAAGAATTGACCTGT
      1260      1270      1280      1290      1300      1310

      1330      1340      1350      1360      1370      1380
mRNApot TTGGAGAAAGCCAGAGCCAGACGTTGCAGACCAATTATGTTGACTACTTGGACGAGCTCG
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Macaque TTGGAGAAAGCCAGAGCCAGACGTTGCAGACCAATTATGTTGACTACTTGGATGAGCTCG
      1320      1330      1340      1350      1360      1370

      1390      1400      1410      1420      1430      1440
mRNApot CCTTAGAGATAGGTGCGAAGCCAGATTTCTGCTCTCTCTTGTTCAAAGATCCTAAACTGG
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Macaque CCTTAGAGATAGGTGCGAAGCCAGATTTCTGCTCTCTCTTGTTCAAAGATCCTAAACTGG
      1380      1390      1400      1410      1420      1430

      1450      1460      1470      1480      1490      1500
mRNApot CTGTGAGACTCTATTTTCGGACCCTGCAACTCCTATNAGTATCGCCTGGTTGGGCTGGGC
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Macaque CTGTGAGACTGTTTTTCGGACCCTGCAACTCCTATCAGTATCGCCTAGCTGGGCTGGGC
      1440      1450      1460      1470      1480      1490

      1510      1520      1530      1540      1550      1560
mRNApot AATGGGAAGGATTCAGAAATGCCATCTTCACCCAGAAGCAAAGAATACTGAAGCCACTCA
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Macaque AATGGGAAGGAGCCAGGAGTGCCATCTTCACCCAGAAACAAAGGATACTGAAGCCACTAA
      1500      1510      1520      1530      1540      1550

      1570      1580      1590      1600      1610      1620
mRNApot AGACTCGGGCCCTGAAGGATTCATCTAATTTCTCAGTTTCTTTCTGTTGAAAATCCTGG
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Macaque AGACTCGGGTCTGAAGGATTCATCTAATTTCCCAGTTTCTTTCTGTTGAAAATCCTGG
      1560      1570      1580      1590      1600      1610

      1630      1640      1650      1660      1670      1680
mRNApot GCCTTCTTGCTGTTGTTGTGGCCTTTTTTGGCCAACTTCAATGGTCCTAGTCAGCATAAT
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Macaque GCCTTGTGCTGTTGTTGTGGCCTTTTTTGGCCAACTTCAATGGTCCTAGTCAGCATAAC
      1620      1630      1640      1650      1660      1670

      1690      1700      1710      1720      1730
mRNApot GCTTTGGGCTTTATTATCTTGTCAGTCACTACCTCCTAAAGAAAAAAAAAAAA
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Macaque ACTTTGGGCTTTATTATCTTGTCAGTCACTACCTCCTAAAGAAAAAAAAAAAAAAAA
      1680      1690      1700      1710      1720      1730

```

Figure 6(suite)

Alignement-ProtFMO2

>gi|1388193 (U59453) flavin-containing monooxygenase form 2 [Macaca mulatta]
Length = 535

Score = 2723 (1271.0 bits), Expect = 0.0, P = 0.0
Identities = 515/535 (96%), Positives = 524/535 (97%)

```

mRNApot:      1 MAKKVAVIGAGVSGLISLKCCVDEGLEPTCFERTEDIGGVWRFKENVEDGRASIYQSVVT 60
                MAKKVAVIGAGVSGLISLKCCVDEGLEPTCFERTEDIGGVWRFKE VEDGRASIYQSVVT
Macaque:      1 MAKKVAVIGAGVSGLISLKCCVDEGLEPTCFERTEDIGGVWRFKEKVEDGRASIYQSVVT 60

mRNApot:     61 NTSKEMSCFSDFPMPEDFPNFLHNSKLLEYFRIFAKKFDLLKYIQFQTTVLSVRKCPDFS 120
                NTSKEMSCFSDFPMPEDFPNFLHNSKLLEYFRIFAKKFDLLKYIQFQTTVLSVRKCPDFS
Macaque:     61 NTSKEMSCFSDFPMPEDFPNFLHNSKLLEYFRIFAKKFDLLKYIQFQTTVLSVRKCPDFS 120

mRNApot:    121 SSGQWKVVTSNGKEQSAVFDAMVCSGHHILPHIPLKSFPGMERFKGOYFHSRQYKHPD 180
                SSGQWKVVTSNGKEQSAVFDAMVC+GHH LPHIPLKSFPG+ERFKGOYFHSRQYKHPD
Macaque:    121 SSGQWKVVTSNGKEQSAVFDAMVCTGHHFLPHIPLKSFPGIERFKGOYFHSRQYKHPD 180

mRNApot:    181 GSEKRLVIGMGNSGSDIAVELSKNAAQVFISTRHGTWVMSRISEDGYPWDSVFHTRFR 240
                G EGKRLVIGMGNSGSDIAVELSK+AAQVFISTRHGTWVMSR+SEDGYPWDSVFHTRFR
Macaque:    181 GFEGKRLVIGMGNSGSDIAVELSKSAAQVFISTRHGTWVMSRVEDGYPWDSVFHTRFR 240

mRNApot:    241 SMLRNVLPRTAVKWMIEQQMNRFNHNENYGLEPQNKYIMKEPVLNDDVPSRLLCGAIKVK 300
                SMLRNVLPRT VKWMIEQQMN+WFNHNENYGLEPQNKYIMKEPVLNDDVPSRLLCGAIKVK
Macaque:    241 SMLRNVLPRTVVKWMIEQQMNQWFNHNENYGLEPQNKYIMKEPVLNDDVPSRLLCGAIKVK 300

mRNApot:    301 STVKELTETS AIFEDGTVEENIDV IIFATGYSFSFPFLED SLVKVENNMVSLYKYIFPAH 360
                STVKELTETS AIFEDGTVEENIDV IIFATGYSFSFPFLED SLVKVENNMVSLYKYIFPAH
Macaque:    301 STVKELTETS AIFEDGTVEENIDV IIFATGYSFSFPFLED SLVKVENNMVSLYKYIFPAH 360

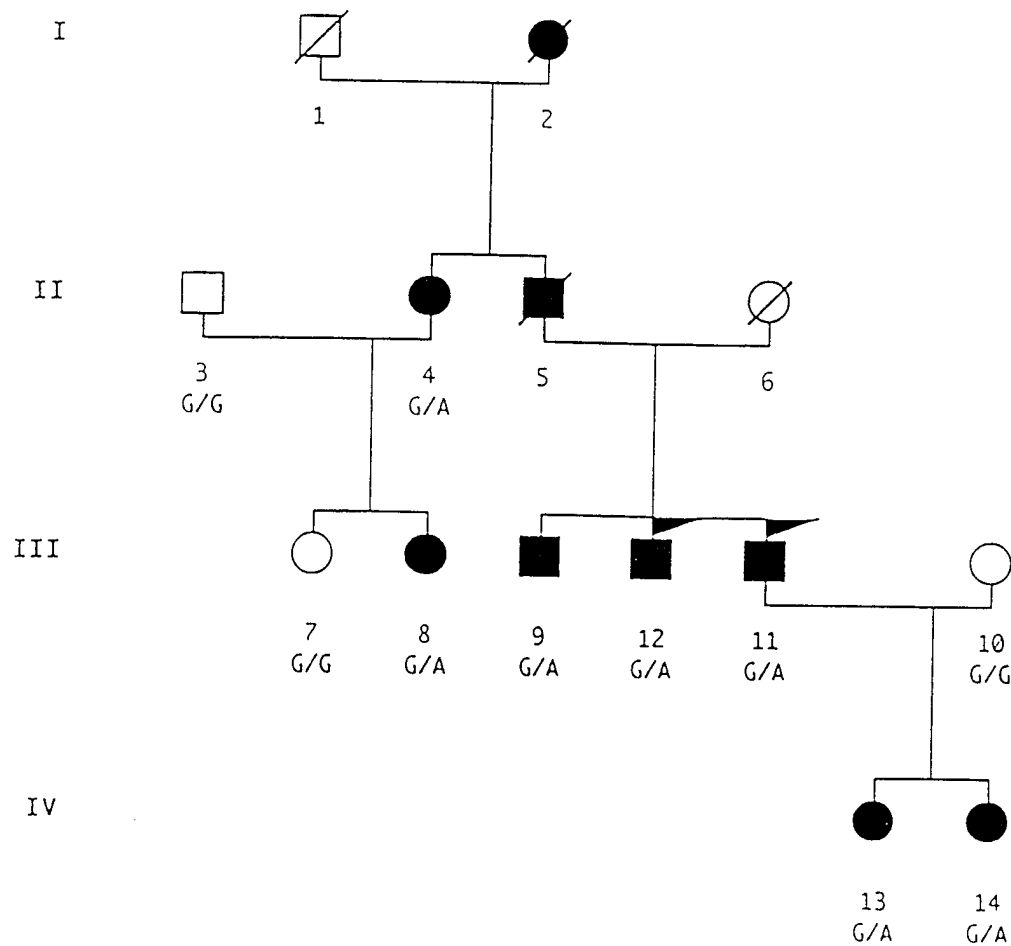
mRNApot:    361 LDKSTLACIGLIQPLGSIFPTAELQARWVTRVFKGLCSLPSERTMMMDIIKRNEKRIDL 420
                L+KST ACIGLIQPLGSIFPTAELQARWVTRVFKGLC LPSERTMMMDIIKRNEKRIDL
Macaque:    361 LEKSTFACIGLIQPLGSIFPTAELQARWVTRVFKGLCHLPSERTMMMDIIKRNEKRIDL 420

mRNApot:    421 GESQSQT LQTNVVDYLDLAL EIGAKPDFCSLLFKDPKLA VRLYFGPCNSYXYRLVGPGQ 480
                GESQSQT LQTNVVDYLDLAL EIGAKPDFCSLLFKDPKLA VRL+FGPCNSY YRL GPGQ
Macaque:    421 GESQSQT LQTNVVDYLDLAL EIGAKPDFCSLLFKDPKLA VRLFFGPCNSYQYRLAGPGQ 480

mRNApot:    481 WEGFRNAIFTQQRILKPLKTRALKDSSNFSVSFLLKILGLLAVVVAFFCQLQWS 535
                WEG R+AIFTQQRILKPLKTR LKDSSNF VSFLLKILGL+AVVVAFFCQLQWS
Macaque:    481 WEGARSAIFTQQRILKPLKTRVLK DSSNFPVSFLLKILGLVAVVVAFFCQLQWS 535

```

Figure 7

27/53
Figure 8

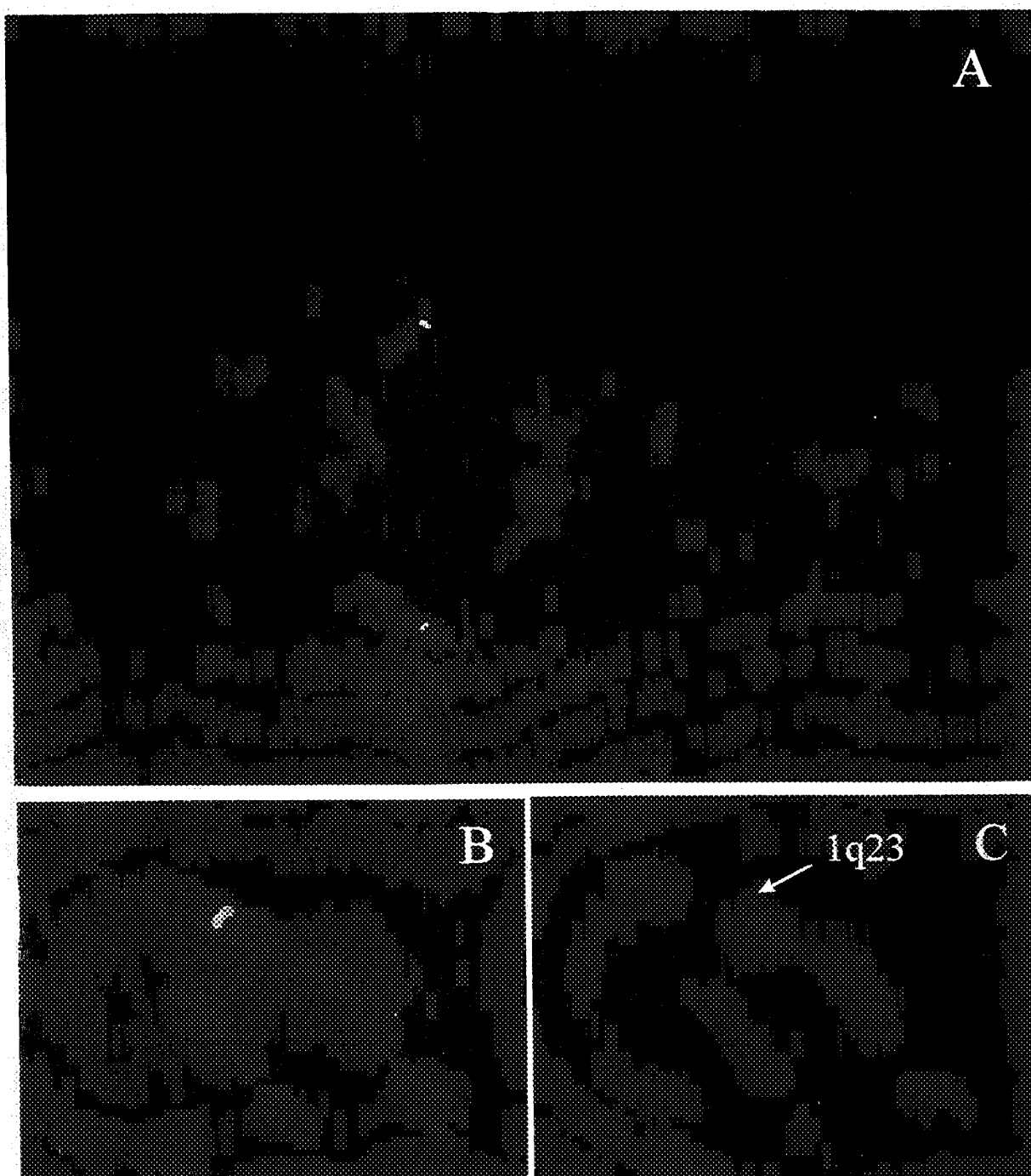


Figure 9

FEUILLE DE REMPLACEMENT (REGLE 26)

29/53

fragment1.FMOx

```

AAACAAGAAA GCTAACTCTC ACCCNTTCTA TTCAATGTAG AACTGAAAGT CCTAGCCAGA
  10         20         30         40         50         60

CAATCAAGCA AGAGAAAAAA AAATCAAATC AGAAATGAAA ATGGTAAATT GCCTTTATTT
  70         80         90        100        110        120

ACAGTTGACA TCATATTTTA TATAGAAAGC CCTAAAGACT TCACCAAAAA CTTTATATAAC
 130        140        150        160        170        180

TAATAAACAC ATTCACTAAA GTTGCAGGAT ACAAATCAA CACTAAAAAA TGGATTGCAT
 190        200        210        220        230        240

TTCCATACAC CAACAACAAA CTATCCAAAA AAGAAATTAA GAAAGAAATT CCATTACAA
 250        260        270        280        290        300

TAGCATCAAA AAGAAAAAAT ACTTAGGAAT ACAATTAGCC AATAAGGTGA AACATCTGGA
 310        320        330        340        350        360

CATTTAAAC TATAGTCTTT AAGGCAATTG AGGAAGACAC AAAAGAATGC AAAGATATTT
 370        380        390        400        410        420

TGTGTTTATT AATTAATATT GTTAAATGT CCATATAATC CAAAGCAATT TACATATTTA
 430        440        450        460        470        480

ATGCAATCCC TTTCAAAATA TAAATAGCAT TTTTCACACA AATATGAAAA AACATTCCTA
 490        500        510        520        530        540

AAATTTATAT GGAACCAAAA ATATCCCAGA TAGCCAAAGC AATCTTGAGC AAAAAGAACA
 550        560        570        580        590        600

AAGCTGGAAT CATCAGACTA CCTGATTTCA AAATATATTA CAAAGCTATA GTATTCAAAA
 610        620        630        640        650        660

CAGTAATGTA CTGGCACCAA AACAGACACA GACCAATGAA AGAGTATAGA GAGCCACGTA
 670        680        690        700        710        720

ATTACCCACA CATATACAAC CAACTAATCT TTGACAAAAG TGCCAAGGAC ACACAATGAG
 730        740        750        760        770        780

AAAAGGATAG TTTCTTCAAT AAATGGTGGT GGAAAAAATT GGATATCCAC ATGCAAAAGA
 790        800        810        820        830        840

ATGAAATTGG ACCCTTAGCT CACACAATAT GCAAAAATTA ACTCAAAATA TATTAAAGAC
 850        860        870        880        890        900

TTAAACTGTA AGTATAACAT GTATATAAAT CATGAAACTG TAAACTACT GAAAAATAAA
 910        920        930        940        950        960

ACATAGAGAC AAAGCTTTTT GAGATTTGTC TGGGCAATGA TTTTTCAGA TATAACCCCA
 970        980        990        1000       1010       1020

AAACCATAGG CAACAAAAGC AAAAAGGNGC CCAATGTGAC TGGCTTCAAA TTAAAGGCT
 1030       1040       1050       1060       1070       1080

TCTATTACAG GAAAGGAAAC AACATAGTGA AGAGACAACC TATAGAATGA CAGAAAAGAT
 1090       1100       1110       1120       1130       1140

TTGCAAAACCA TGCATGTGAT ACGGGGTAA TATCAAATA TATAAGAAAC GTAAACATGT
 1150       1160       1170       1180       1190       1200

CAACAGCAAG AAAACAAATA ACCGGATTTT AAAATGGGTA AAGGACCTGA ATAAACATCT
 1210       1220       1230       1240       1250       1260

AAAAAATAAA TGGCCATATG AAAAAATGCC CAACATCACT AATCATCAGG AAAATGCTAA
 1270       1280       1290       1300       1310       1320

TCAAAACAAC AAGATATCAC CTCACACCTA TTAGGATGGC TATTATCAAA AAGATGAAAG
 1330       1340       1350       1360       1370       1380

ATAACAAGTG TTGGTGAGGA TGTGGAGTAA AGGGAATCCT TGAATGCTGT TGATGGGAAT
 1390       1400       1410       1420       1430       1440

GTAAATTAGA ATAGCCATTA TGGGAAACAG TATGAAGGTT CCTCAAAAAA GTAGAATAGA
 1450       1460       1470       1480       1490       1500

ACTACCATAT GATTCTGCAA TCCTACCTCT GGGTATATAT CCACAGGACA TGAAATCAGT
 1510       1520       1530       1540       1550       1560

```

Figure 10

30/53

fragment1.FMOx

```

AAGTCAAAGA CATGTCTGCA CTCCCATGTT CACTGCAAAA TTATTCACAA TGACCAAAAT
1570      1580      1590      1600      1610      1620
ATAGAATCAA CCTAAGCATT TATCAACAAA TGGATAAAGA AAATGTGGCA CTTAACTACA
1630      1640      1650      1660      1670      1680
GTGGAATATT ATTCAGGCTT AAAAATGGAA AACATCTTGT TGCAACAATA TGGATGAACC
1690      1700      1710      1720      1730      1740
TGACAGACAT TTTGCTATGT GAAATTAACC TAGGCCCAGA AAGACAAATA CTGATCATTC
1750      1760      1770      1780      1790      1800
TCACTTACGT GCAGAATCTA CAAAAGATAA ACTCACAGAA GCAAAGAGTA GTAGGGTGAT
1810      1820      1830      1840      1850      1860
TTCCAGGGGC TTGAGGGTGG GGGAAATTGA GTAATCTTGT TCAAATGCTA CAAAGTTTCA
1870      1880      1890      1900      1910      1920
GTCATGCAGG ATGAATATGT TCTGGAGATG TAATGTACAA TATGGTGAAT ATGGTTAAAA
1930      1940      1950      1960      1970      1980
TACTGTATGT TATGCTTCAA ATTTGCTAAA AGAGTAGATC TTAAATGTTT CACTCAGAAA
1990      2000      2010      2020      2030      2040
ATAATGGTAA CTATGTGAGG TGATAAGTAT GTTAAATAGC TTATCATTTT ACAATGGGGA
2050      2060      2070      2080      2090      2100
GATATATATA TATACGATGT TTTGTGCCTT AAATATACGC AGTTTTTATT TGTACATAC
2110      2120      2130      2140      2150      2160
CTTACCTCAT GGTTCGATT CTGCACTCTG CCCAACTTTC CGTGTGCACT CTCTTGGGTA
2170      2180      2190      2200      2210      2220
TGTGCTCTTG CTCCCATGGT CCCTGTAGCT GACAACACTT TCCACCCATT TCAAACCACA
2230      2240      2250      2260      2270      2280
CATTTTCTTT CATTTCTCAA AGTCCTTTCT CAAACACTGA TGAAAATGCA AATCTTCTCT
2290      2300      2310      2320      2330      2340
CAGTCCCCTC CCACAGAGTA AATTTCTCAC CCTTCTAGGT TTAAATAGCC CCTTTTAAGA
2350      2360      2370      2380      2390      2400
CTCTGTTACT ATGTGAATCA TGGTCAAAAC TGTAGCATAG TTAACAATGT ATGTGTCTTC
2410      2420      2430      2440      2450      2460
CTCTCATCTA TCTGGCCTAA TGCTGTGACT GATACCTATT AAATATTTGG TACATAATAA
2470      2480      2490      2500      2510      2520
TTCTTTTCTT TTTCTTTTCT TTTCTTTTCT CTTAAGGGGA GGCAGCACAT CTTACAGTCT
2530      2540      2550      2560      2570      2580
CTGCCACATC TTAGTTGTAT TAATTGTGCC AGGTTAACCT GCTAAGCCCC AGTCTTTCTT
2590      2600      2610      2620      2630      2640
CAGAAAAATG GCCACAGTAA TACTTGCCTT ATAAGGTACT TTTGAAAACA AATAACATAA
2650      2660      2670      2680      2690      2700
TGTGTCTAAA GTGCTTAGTT TAGTGCCTGG CACCCAGTGA ATGCACATTC TAAAGTGATT
2710      2720      2730      2740      2750      2760
ATTTTGTAGT GTATTTTCCC CCATCATTTT CTGACTGTGA GGCAGTCTGA AGCATTTTAC
2770      2780      2790      2800      2810      2820
ATACCATGTC TCAGAGTCTG CACTAGAACT TAGCTTCCTA GGAGACGTTA CTCTGCTCCT
2830      2840      2850      2860      2870      2880
TTTTGTGCCC ACTGTATCTG TAACATTGCA AACAATGCCT GGCATGCAGT AAATGCTCGA
2890      2900      2910      2920      2930      2940
TTAATACTTG TTGAATAACC CTCTGAAGGT TCCATACTCT CTAAAGTAAG AGTATCTTCT
2950      2960      2970      2980      2990      3000
TCATACTCTC TAAAGTATGT CTTCTCACTA TTCCCATTTA TAGGGTATTG GAAAAAACAT
3010      3020      3030      3040      3050      3060
AGCATCAAGT CATCAATTTT TGCTGGCTTA GCAAATTTGA ACAGTTCAGT TTTTCATTAG
3070      3080      3090      3100      3110      3120

```

Figure 10(suite)

31/53

fragment1.FMOx

```

TATATTGAGA ATAATAATAC CTGCCTCACA AGGTTAATGG AAGAAAATAA ATGAAAAGCA
  3130      3140      3150      3160      3170      3180

CTTTGAAATA CGTGTACGCA TTCTGAAGAT GTTCATTAAT CTCAGTCTAT TCTTAATCCT
  3190      3200      3210      3220      3230      3240

GTTCTTGGAG TGGCTAGTGT GAACCATTAG GGGCTGCAAG CTCAGTGGTA TGACCTATAG
  3250      3260      3270      3280      3290      3300

AGCCCTGAGC TTTAAGAGAA AAGCATGGAA CTTTAAGTAC AAGGAGACAT GGAAAGGCCA
  3310      3320      3330      3340      3350      3360

CAGAGAGGTA GATTCCTCAGC TGAGGGGGAG GTAGAGAAGC TCTGGAAATG ATCACCTTTA
  3370      3380      3390      3400      3410      3420

TCCCTTGCCT TAACATATCT CTTCTCCTG GGTCTCTCCAT AGTACTCAGG GTCAGCCTGA
  3430      3440      3450      3460      3470      3480

AGATGGTTCA CCACTAGAGA ACTGCTGGCG CTGAAGAAGA AGAAGACACT GTGGACAATC
  3490      3500      3510      3520      3530      3540

TTCTCCCTTT GTTTCATGAT CTTGCTCACT TTGGCCTATC AAATAACTCC AGGGACTTCA
  3550      3560      3570      3580      3590      3600

TCTTTTAGTT GATTAAAGTAC CTGGGATACC ACCTTCTTAT AGCTCTAAAG ACACTGCCAA
  3610      3620      3630      3640      3650      3660

GCTATGAGAA TGCAGGGGGA GGCAGGAGAT CCACAAGGTC TGATTTTCCT GATCATTTTT
  3670      3680      3690      3700      3710      3720

GAGTAACAGG ACTTCTCTGT ATCTCATTAA ATCACAAGTG CAGTCTCAAT CCGGAAAATA
  3730      3740      3750      3760      3770      3780

GATCCCATCA TATGTGATAA TCAAAACCTT GGACTTTACT TTCAGCTGTA ATCTTTTCAG
  3790      3800      3810      3820      3830      3840

TATTACCTTC AGCTGAACCT GCAGCTAGGA GAAGCCTGAG GTCCTGGAAG GCAAGCATGA
  3850      3860      3870      3880      3890      3900

GGCTTTTCTC CTATATCGCC TTATGCTTCC ACTCCTCCTC CCCATGATGC TGATGGAAAA
  3910      3920      3930      3940      3950      3960

GTTTCCAACC CTTTCAGAAA ATTATATTTT CTGTCTTTGA AAAGCTGACA TTTTACAAAT
  3970      3980      3990      4000      4010      4020

CTCTTAATTT CCTTAATACA CTGAGGTGTA TGCAGGAAAT TAAGAGATTT GTAAAATGTT
  4030      4040      4050      4060      4070      4080

GAAGGGATAT TAGTAATTAC TTAATCCAAT TTGCTTATCT TACAGGGGAA GGATCTAAAT
  4090      4100      4110      4120      4130      4140

TCCTGCAAGA ATTATCCCAA AATATACAGC CAAAGAGATG AATCACTGAG TTTTAACTA
  4150      4160      4170      4180      4190      4200

TAATATTAGA TAACAGTATG AATATTTTAA AAATTATATA GAAGACAGAA CTGGCTTTCA
  4210      4220      4230      4240      4250      4260

TTAATATCA TACGTATTAC AAGAATTGGA AAGCTAAAAA ACCACATTTT TAAGACTCTC
  4270      4280      4290      4300      4310      4320

TTGCAGTTAG GGATTGGATG CAATTAGTTT TGCCAGTGAG ACAGACATCT AGGCTCAGAA
  4330      4340      4350      4360      4370      4380

GGCAAAGGGA AGCAAAAGCT AAATTTTCGT GACAATGCTG ATGGGCAAAAT GTATTCCAAC
  4390      4400      4410      4420      4430      4440

AGGTGTGCTT ATATCAGCTG AACTCCATGT TCCACCTTTT AGCCCCCAA TTTGTATGTT
  4450      4460      4470      4480      4490      4500

GCAGGCATCT GTGGTGATTG AAGCAGTTTC CTGCCCTCTG GATAGTAGCT CTAGTGGTGT
  4510      4520      4530      4540      4550      4560

GACCTCTCCA GTCCTGCAGC AACCTCCTAG TTCCTGCTCT TTCAGCTCTT TGACCTTTTG
  4570      4580      4590      4600      4610      4620

CAAGCACCTA ATTCCCTGTA GTATATACCT TTCTTCATGA TATATAGTGT TTTTATCTC
  4630      4640      4650      4660      4670      4680

```

Figure 10(suite)

32/53

fragment1.FMOx

CTGCACTAAA	TCATGAGCAT	ATGCATATAA	ATCATAATAT	GAAATCTTAA	AAACAGAAGT
4690	4700	4710	4720	4730	4740
ACTTTTGCTG	AGGCATTAAG	CATATAATCA	GTCAGCAGGT	CCCCAAACAT	CTAATTCCTG
4750	4760	4770	4780	4790	4800
AATATCTCAT	ATATCCTGTC	TCCATTATCC	ATTCTCTCTAA	TGCTACTCTA	ATTTAAGTCC
4810	4820	4830	4840	4850	4860
TCAGTCTCTC	TGGCCTAGAT	TGTTGAAATA	ACATCCTGGG	TTTTTGGTCT	CCTTGATTCT
4870	4880	4890	4900	4910	4920
AGTCACCATC	CTCTCTAGCC	TCCAGGTGAA	TCTGATCTTG	TCTGATGTTG	TCACTTCCTT
4930	4940	4950	4960	4970	4980
GTTCAAAATT	CTCGAATGGA	CAACCGTAAT	CCAGAAGGTA	GATCCCAAAC	CTGTGATTGT
4990	5000	5010	5020	5030	5040
GGCACTTCAG	TATCCTTCAT	AACCTATGTC	CTGCATGTTT	AACCCATATT	TTGCTATTCC
5050	5060	5070	5080	5090	5100
CATCACTTAT	GGTCCAGCAA	AACTGAACATA	ATTGTAGTTC	CCCCATCACG	TGTTCTTACT
5110	5120	5130	5140	5150	5160
TTTCTATGCA	TTTTCACATA	TTTTTCTCTC	TGCCTTTTTT	CTATTCTTTG	TCCCTTATCT
5170	5180	5190	5200	5210	5220
GTCTGGAAAA	CATCTATTCT	TCCTTCAAGA	CTCAGCTGTC	TTCTCACACT	CCTTGAAGCC
5230	5240	5250	5260	5270	5280
TCTCTTTTCT	CCTCCAAGTG	GACCTAGATT	TTTCTTCCTA	CATGCTAGCA	CTACACTGAA
5290	5300	5310	5320	5330	5340
CCATACTTCC	ACTGTGACAT	TTATCATCTC	CCTCAACACT	AGACTTCATG	GTTCCAGATG
5350	5360	5370	5380	5390	5400
GAAAGCACTG	TGTCTTCTCA	CCTTTGAATC	CCCCAAAAGA	CTATTATAAT	GCATGACATA
5410	5420	5430	5440	5450	5460
TAGTAGGCTG	TCAGTACAGT	GAAAGGAATG	GCCAGAGGAA	GGAAAGGAGG	GAAACAGAAG
5470	5480	5490	5500	5510	5520
CAGAAAGGAC	AGGTATAGAA	GCCGGAGGGA	GCCAGAGACA	AGGTTTCAGAG	ACCACAATTC
5530	5540	5550	5560	5570	5580
TGTCTTTTGA	GTTCACTAGT	TTTACAAGCT	CATCTATAAG	CGTTAGTTCA	GCAACTCAGA
5590	5600	5610	5620	5630	5640
TCAGGCCCTA	AGTTTCCAGA	AATTTGAGCT	ACTTTTCACT	GTTGGCACAA	CAAAACGTTT
5650	5660	5670	5680	5690	5700
CATTATAGTC	CAGGTGCATA	GCCTTTGTTT	ATATATTCTA	TATTTCCAAA	GCAAACATAA
5710	5720	5730	5740	5750	5760
ATGAAAGAAT	CATTGTTCCC	CTAATCTCCC	AGGAGTTTCA	CCTTACAGCT	CCAGTGGCCA
5770	5780	5790	5800	5810	5820
TGGCAGTCAC	TGTTTTATAT	TTTTTGTAAC	AAGAACCAAA	GACTTCATTC	TTCCTTTTTC
5830	5840	5850	5860	5870	5880
CTACCCCTTT	CTTTTACTT	CACCCATGCC	TCCCCTGTTT	TTCTCTTATC	CCTACCACAC
5890	5900	5910	5920	5930	5940
TCGTCTTTCT	CTTTCAGATT	TTACTATGGC	TCTATACCAT	TAAAAATACA	AGAAAAAATA
5950	5960	5970	5980	5990	6000
GGAATTTTAC	TTTAAGAATA	ACTCTCCCCC	CTTCCCCAGT	TTTCACATCA	AAAGACATTG
6010	6020	6030	6040	6050	6060
TTAAATGCCA	TTCTCTTCCA	CATTTTCGAGA	ACTGCTGATT	CTCTGGGGAG	AGAAAGGTGA
6070	6080	6090	6100	6110	6120
TTGCTTAAGA	GGTGAAGTCC	CTTAGAGCAT	TCAAAATGAG	GAGTGATTCT	GTACAGAGGA
6130	6140	6150	6160	6170	6180
TATCATGCAG	CAGGCTGGAT	GTCTAGTTCC	AATTCCTTTA	TTTGTACCT	CTGAGACCTT
6190	6200	6210	6220	6230	6240

Figure 10(suite)

33/53

fragment1.FMOx

```

GAAGAAGTAG TTTCTAGTCT CAGCATACCA AAGCGTCATC TGCAATTGAG AGCATTGGAT
6250      6260      6270      6280      6290      6300

TGATGATCTT CAAGGTCCTT CCTGCTCTAG CATTCACTGA ATCTGCTATT TTTGACATAT
6310      6320      6330      6340      6350      6360

TGAATAATCA GAAGCAGCCA GTTTTAGAAT CTTATTATAG CAAAAGTGGT AAAAATAATG
6370      6380      6390      6400      6410      6420

AGCATATACT ATCAATGTGC ATCTATGTCT TCTTATGTTT GAGTGAGGAT CCTGATACAT
6430      6440      6450      6460      6470      6480

AAACCTTGGC TGATAATTTT TACTGAAAAA AATCGTAACT ATTAAAGACA CTCTTCTGAA
6490      6500      6510      6520      6530      6540

GATGTTCTCT CCAGACTCTG CTACAGGCAA TCATGAGCAA GAGGGTTGGC ATCATCGGAG
6550      6560      6570      6580      6590      6600

CTGGAGTCAG TGGCTTGGCT GCCATATGGT GCTGTCTGGA GGAGGGGGCTG GAGCCCACTT
6610      6620      6630      6640      6650      6660

GCTTTGAAAG GAGCGATGAT GTTGGAGGCC TGTGGAAATT CTCAGTGAGT GGCACATCAT
6670      6680      6690      6700      6710      6720

TAGAACACCA GTGGAAGGAG ATGGATTCCA ATGCAAATCA AATCTGATCA GTTCTAATTC
6730      6740      6750      6760      6770      6780

AGATTTAGAA GGCAGATCAC AAAAGCTCCA AATCTGGAAA GTAAAATCTT ACCTCTCCAA
6790      6800      6810      6820      6830      6840

TCATACTAAT NGCCCCAAAA AACTATTTCA TACCAGCAAA ATTTGTCTCTG AAAAGGACAT
6850      6860      6870      6880      6890      6900

TTTCAGCTCA TTAAACATCA TCACCTGCAT GGTGAAATCC AGATCTCCAA GCTGTAAAGG
6910      6920      6930      6940      6950      6960

GCACTAATGT TGGTAATTAT TCAAAAATAT ACCATGGGCT TCCCAGGTAA GTGAAACAAT
6970      6980      6990      7000      7010      7020

TCTATTCTTT ATTGCTCTTA AATGCCAGGA ACACGACTAG AAAAGAGACA AACAAACCTG
7030      7040      7050      7060      7070      7080

GACTGAGATC CTGAGGTCAG AAGTCCTGAG TTCTAATTTT AACTTGTAGG TTTTCTAGGC
7090      7100      7110      7120      7130      7140

AGATAAGATT TCAGTCCAGT TGCTTTTGT TCCCTGGACC TCAAATGCTC ATTTGTCAAA
7150      7160      7170      7180      7190      7200

TGCAGAGGAT ATGATTCTAT AATTAACCTA TGTCTATTGG GCAGATAGAG ATTATTATAG
7210      7220      7230      7240      7250      7260

ATGATGATTG TGTGTGCGGC TGTGAATAG CCTATCAGCT CCAAATCCAG AGGGAAAAAT
7270      7280      7290      7300      7310      7320

TATGGTCTTT GCCATTGGG CTCATTGTAG AAATAATATA ATTAGGAAAT AGTCCTTGTA
7330      7340      7350      7360      7370      7380

AACACATTTT TTTTAAATT TCAAAGCCAA GTTTGGAGAA ACTTCTAGTT CTTCTGTCCT
7390      7400      7410      7420      7430      7440

GGATTTCCTA GCCATTGTAA TCAGTTGTCG ATGATACATA TTTGGCTTGA AAACATATTC
7450      7460      7470      7480      7490      7500

ACATCATTCA TATTGTAACCT ACTTCCTGTC CTGGTCTCAG TTACTGCTCT GCCTGCGCCA
7510      7520      7530      7540      7550      7560

ATAGCCTCCT CCAATAGAGT ATATCAGTGC TAACTTAGAA CACATTTTTA TTCTTCTCCA
7570      7580      7590      7600      7610      7620

AGCTTTTTTT AAAAAAATT GTGTTTTGT AACCTGAAA GCACTCCATG AGATATAAGG
7630      7640      7650      7660      7670      7680

TCATTAATTT TTATTCCCA GTAGGGGGTA ATCAAGAGTT AATATTTTTC AAGAATTTAA
7690      7700      7710      7720      7730      7740

TTTTCCCTAT TTACATTTGC TCAGGGAAAT GTGGACAGCT TAGAGTAAAT CATAAAATGG
7750      7760      7770      7780      7790      7800

```

Figure 10(suite)

34/53

fragment1.FMOx

```

CTTCTACCA TCTCCCTAGT AACAAATAAA TGATGCTTGA GCATCTATTC TGGTAGTTTG
7810      7820      7830      7840      7850      7860

TGCTAAGTAC TGGGATGACA AATATGGAAT ATAATCACTC CTTGTAAATG GTTCCATTTC
7870      7880      7890      7900      7910      7920

ATTTGATTAA GCAAGCCATA ATATAATTCC GTAATCCTTT GATAGCAAAT GGGCAAAAAC
7930      7940      7950      7960      7970      7980

TCATTTGATA GCCGAACCTC TTCTGAAATC GTAAGGTAA ATACCGTGAA TTGGATCAAC
7990      8000      8010      8020      8030      8040

ATGAAGCTAA GTCTCACCTT CTGTTGCACG GCAGAAATTT TATTGCATTT GACAGATTGC
8050      8060      8070      8080      8090      8100

TGCCCCAGAT CTCCTAGGTA GTATTATGGA GCAAAATCCA AAAATGTACA CATTCCAAAA
8110      8120      8130      8140      8150      8160

TATATCTGGC CCTAAGACTT TTAATAAAG AGATTATATA ACTACAACAA CAAGATAGAC
8170      8180      8190      8200      8210      8220

CTTGTCACCA TCAATTCAAT GGACGAGTGC CTCGAGCGTT TAGAAGAGGG TGGACTACGG
8230      8240      8250      8260      8270      8280

AAATCTTAGT AGGTCAAAGA AAACCTCCCA CAGGGATGAC ACTTAGCCTT GAAGGATAAC
8290      8300      8310      8320      8330      8340

CCCAGACAAG CAAAATAGAG GACCACCTGT GACACAATC CTAGAGAGTG CATTTCCTAA
8350      8360      8370      8380      8390      8400

TAAAGTCTGC GAATGGCACT CCATAGGCCT ATGCAGTCAT TGGCAGTGTG CCAGACCAG
8410      8420      8430      8440      8450      8460

GTTAAGAGAG ACCAACAATC CATGAAAGGC ACAGAAAAGG CAATGAACAT GGTGTGTGCA
8470      8480      8490      8500      8510      8520

GAGAGGGATC CATGAGTTAT CCAATATAGC CAGATCAGAA AGTTTACTTA AGGAAGCAAT
8530      8540      8550      8560      8570      8580

AATATGATAC AAAGATCAGT AAGATTCAAA GTTGGATTCT GAGTTATCCA CAAGAGGAAA
8590      8600      8610      8620      8630      8640

TTCTTTCTTT CCATAAGGTC ATGTCTATAA GCAAAATTCT ACTCAAAGTC CTGGTGAGGA
8650      8660      8670      8680      8690      8700

TATGGACCCA TACAAATACT CAAAACCTTA GCCTCCTCCA CATACCCAG CCCTTCCTTC
8710      8720      8730      8740      8750      8760

TTTTCTTAGA AAAGTTGCTT GGCACAATAT ATAATCAGAG AGGGATTTTT TTTATGTGTT
8770      8780      8790      8800      8810      8820

ACATAAGACT TTATCTGTGN AAGNCTTTTT TNCAGAAGGT GTTCTAGCAG ACAGAAACCT
8830      8840      8850      8860      8870      8880

GGTAATTCTG AACTTTTCAC TATTTGCTTT TTCTGAGAAA TGAAAACCAA ATGGGATTTA
8890      8900      8910      8920      8930      8940

AATACTAGCA GGCTGAATGT GTGTTTAAAG TTTCATCCAC TCCTAAATAG GGCCTCGTGT
8950      8960      8970      8980      8990      9000

CCTCAAAAGA TTTCACTACT GCTGTAATAA GAAGTTGCTC AACAGCCAGG TCGGCTGGCT
9010      9020      9030      9040      9050      9060

CATGCCTATA ATCCCAGCAC TTTGGGAGGC CAAAGCGGGT GGATCACGAG AGGTCAGGAG
9070      9080      9090      9100      9110      9120

TTCAAGATCA GCCTGGCCAA CACAGAAAAA CCCCATCTCT ACTAAAAACA CAAAATTAG
9130      9140      9150      9160      9170      9180

CCAGGTGTGG TGSTGGGTGC CTGTAATCCC AGCTACTCAG GAGGCTGAGG CAGGAGAATC
9190      9200      9210      9220      9230      9240

TCTTGAACCC AGGAGGCAGA GGTTCAGTGC ACCTCAGATC ATGCCACTGC ACTCCAGCCT
9250      9260      9270      9280      9290      9300

GGGCAATAGA GAAAGACTCC ATTAATAAAA AAAAAAATGC TTA
9310      9320      9330      9340

```

Figure 10(suite)

35/53

fragment2.FMOx

```

aaaaaaatgc ttaccaatag gttagtagca ttttgattgc aaaagctgaa aacaggacta
   10         20         30         40         50         60
ttttgaactt tttccactc atttattcct ttgttcattc aatgaatata tactgtgtac
   70         80         90        100        110        120
tttatgtgta ggggtactata ttaagcataa gctgcagata agaggccagc cagcacttta
  130        140        150        160        170        180
aaagccgtga gaaaacaagt atcagaataa ctataagtga ctatataatt agggcaataa
  190        200        210        220        230        240
ggataatggg accttagtaa aactaaagat gatttggcag tagctgagag ggaaggtaaa
  250        260        270        280        290        300
gaaagccatg acaaagtga aggcaacttt tgagcatatt tcaagggcat atttagacaa
  310        320        330        340        350        360
ggagatatgg gactcataag cagagctgga ataggaaaga agatcaaggt aaactgctta
  370        380        390        400        410        420
gatgcatgta caacattctg aaattaacct ctgactttgc cctcaagtta cttatgttct
  430        440        450        460        470        480
cgtgggaaag atgagagatg aacacgggta tcatccaaga cagatgggtgc ccacagctgc
  490        500        510        520        530        540
ttagatctct gggtccaggg taaagctccc tcagctagag gcagagtcaa agttgaattt
  550        560        570        580        590        600
cctccttact ggctcaaacc acacctcata ttgaaataat aaaaatgcat gctccctgga
  610        620        630        640        650        660
gcaactgact tgttatctaa tacatttgct tttttgtgtt tgtttggaaga acagtctttt
  670        680        690        700        710        720
cgaaaaaatt ccaaggagct gtagtgtaca tactctctct tcctggtgtt ataattggct
  730        740        750        760        770        780
gagggtcaagg ggcaaaaaag cagagattca ttcaagatgg aaatattcca aggccttagc
  790        800        810        820        830        840
atctgtttcc cagaacagag tcttacattc ttttaaccagg ctccatccca cagttcagcc
  850        860        870        880        890        900
ctgcctcctt tcaacaggca gctgaaaaaa cctccttccc acctctcctt ctcacaacca
  910        920        930        940        950        960
tcagtagaag gcgctagctg tgggtgaaag ggaagcactc agcctgccaa actgctggac
  970        980        990       1000       1010       1020
atgagccttc accctttttc tgacctccac aaaaatttta aaaagttaa attcctgtgc
 1030       1040       1050       1060       1070       1080
ttccacgctt atgagaaata cagcaaccat gaatagagga agattatggt ttcaacttga
 1090       1100       1110       1120       1130       1140
gaaaaaatac tgaggctttg ggcagccccc cacttcccca cggggacaca atcctctcaa
 1150       1160       1170       1180       1190       1200
ccctttccag cactttttgt ttccctcttc cagaggtcat ctggtgtgag agggagatac
 1210       1220       1230       1240       1250       1260

```

Figure 40 (suite)

fragment2.FMOx

```

acatcttgaa tccagcagca acgtgacatt ccattctctt ccccccattg cacaagagtc
1270      1280      1290      1300      1310      1320

ccttccggac ctccgggaagc agaagctgcc agctctgaaa tgtattttca aggcagcaca
1330      1340      1350      1360      1370      1380

ttgtgtgcac ttttacccta ccctcacaac tgagaggaaa tgtttatttt caatttagct
1390      1400      1410      1420      1430      1440

tttgactgct tctaaaaaat aagccacttt tcaattacac agaggcttta aaatgaagtg
1450      1460      1470      1480      1490      1500

ccaagattta acacatgttc taagggtctct ggtttcctgt gtttcttttg tgaggagtga
1510      1520      1530      1540      1550      1560

agtccagcaa ctggtgagcc aaagaatagg attcatttac aacagagcag tggttctcga
1570      1580      1590      1600      1610      1620

agtgtggttc ctaaaccagc cacatcagca tcaccaggaa cttgatagaa atgcaaaaca
1630      1640      1650      1660      1670      1680

ccccagact ccacccagca cagattgaat ccgaaattct aagaataggg cccaagaatc
1690      1700      1710      1720      1730      1740

tacggtctag ggagcttcca ggcgattctc attacgcaa agctgggaaa ccactgcaat
1750      1760      1770      1780      1790      1800

attgggttgt tgccagtga ggttttgcta aactccaaaa gcaaataaat aggctagaag
1810      1820      1830      1840      1850      1860

tcagagcctc ttctagacag ttttgttttt tgtttttttt ttaacctgag tataagatca
1870      1880      1890      1900      1910      1920

gaaccagtgg tggcacagga gaaagcaaaa accactaagt ggctataaag acagagctaa
1930      1940      1950      1960      1970      1980

cactgagggg aattacagta agaggattca catggaaaag gctccagttc tgtgccaggt
1990      2000      2010      2020      2030      2040

tacgcgaagg gctttccatt ccttatctta ctgagagctt ttaatttttg tttacgcttt
2050      2060      2070      2080      2090      2100

taaacaatgaa aagggtttta gtcaaccaag aattgaacca ctgtgttcac tgaagggaac
2110      2120      2130      2140      2150      2160

-acaattcttg gctttctctt taagctttct tattctccct aggaccacac agaagaaggc
2170      2180      2190      2200      2210      2220

agagccagca tttaccagtc tgtattcaca aactcttcca aagaaatgat gtgctttcca
2230      2240      2250      2260      2270      2280

gacttccctt atccggatga ttacccaaac tatatacacc acagcaagct ccaggaatat
2290      2300      2310      2320      2330      2340

ataaagacat atgctcaaaa gaaggaaact ttaagatata tatagtttga ggtaggggtc
2350      2360      2370      2380      2390      2400

tcataacttg tactgttgaa attaagatat gtgtgggtta gagaaaaagg aggcagcaaa
2410      2420      2430      2440      2450      2460

ctattataaa aattagagcc aaatgttttg gcacctcagt aatcaaatgt tgggtctgat
2470      2480      2490      2500      2510      2520

```

Figure 10 (suite)

37/53

fragment2.FMOx

```

tataaagcat tcatgcattg attttttctc tcctagactt actagttcac tagtctctga
2530      2540      2550      2560      2570      2580
gagcttttcag actaccttag aaaatggagg cagctagccc atcattgtcc actttccacc
2590      2600      2610      2620      2630      2640
ctcatgctct gatgttttgg aaataatcca aaatgcttta gtatatatta ggaattttgt
2650      2660      2670      2680      2690      2700
cagttcaatg ccaatgagtt gtggttcaaa aaaccagagc atttggtagg gtttctccca
2710      2720      2730      2740      2750      2760
ttacattatg aaaagggttaa caacttaaat gggaaatata gtcattgccc ccatctttac
2770      2780      2790      2800      2810      2820
ccactcagtt cattagtttt tttattaaaa aggtgagatt tcagcattgt ttctgcgaga
2830      2840      2850      2860      2870      2880
ataatgtttt acattttattt gggactcttt attgagcatt tctgtctgta tgtttggaac
2890      2900      2910      2920      2930      2940
tcttaacctc aattaactgc tgctaaatgc xgaacacttg catatagtgg gaaaaacaat
2950      2960      2970      2980      2990      3000
cagcaaaatt atgaaccatg gtgatattta catcattatt ttacctggag tagccccaac
3010      3020      3030      3040      3050      3060
tgtatagtta aaataaaatt ttccaatagt cattttattc cattcattca ttacattcat
3070      3080      3090      3100      3110      3120
ttgcttccat tatggtgtta atatcaacaa acattaatga agttccctatt gtgtgcttgc
3130      3140      3150      3160      3170      3180
attgtgctat gtgttatatg taaaagaaaa agaggtctaa gacttagctc tcaagaagtc
3190      3200      3210      3220      3230      3240
atttcaaaat aaatatgtaa agagtaagta aaaagattcc agtaacaatt tcaatcaaag
3250      3260      3270      3280      3290      3300
agaaaaattt ttaaagctct ttatgatttg tttataaata aaacaatgct atggagatca
3310      3320      3330      3340      3350      3360
tgaagcaaga ggcaacactt tgggggaagg tattttctag aggaggtaaa atttagttgt
3370      3380      3390      3400      3410      3420
atttagtagg tgttttagat aaatgagtgg catgagtaaa attagagagg tgggaaaatg
3430      3440      3450      3460      3470      3480
ccctgctcat ttggagaaca gtgggcaaac caagttggtt aggagggaga tatatatgct
3490      3500      3510      3520      3530      3540
aggatgagat atggccacat atatcagtaa actagtgtgt actgtgactt tgaaaaatag
3550      3560      3570      3580      3590      3600
aggattatth tgcaaccatg taaaagaagt ccaaagaagg gacatccaga gcttatgtga
3610      3620      3630      3640      3650      3660
tggcaccaaa gttatcaaag attcagcttc acccatctta gcaccgtggc ctacatcatg
3670      3680      3690      3700      3710      3720
acgtttgcct tgtggtgcaa aacagttgct gaagcttgag ccgtcacatc tgccttctag
3730      3740      3750      3760      3770      3780
gcaaaaaaaaa aaaa
3790

```

Figure 10 (suite)

fragment3.FMOx

```

tttttttttt tttttttgag acggagtctc gctctgtcgc ccaggcggga gtgcagtggc
  10      20      30      40      50      60
gcaatctcgg ctcaactgcaa cctccgcctc ccgggttcac accattctcc tgcctcagcc
  70      80      90     100     110     120
tcctgagtag ctgggactac agggcggcgc caccacgccc ggctaatttt ttgtattttt
 130     140     150     160     170     180
agtagagacg ggggttcacca tgctagccag gatggtcttg atctcctgac ctcgtcatcc
 190     200     210     220     230     240
acccgcctcg gcctcccaaa gtgctgggat tacaggcata agccaccgcg cctggccgta
 250     260     270     280     290     300
aagttgctat atttctaaga taagagtatt tatgcagagc aaaagagatg ccaacgatca
 310     320     330     340     350     360
aaccttgaga tattcccata cttattgagt agatggaaga tgaggtcaga aaaggaggaa
 370     380     390     400     410     420
gccatgtcag tagagggtag ccataagaaa ataacacaga tttgttatat gacatcattc
 430     440     450     460     470     480
acaaaaatat tcagtgtgat ttacccttaa atcaactaac ttgatgtcaa aaagtaaatg
 490     500     510     520     530     540
tactccagtg agtaattttt cttgtgagat tcaaagactc actgaagatt cactgtgact
 550     560     570     580     590     600
ccaattttac tatctttcta tacatttctg aatgaccaag agagctcgtc acaattattt
 610     620     630     640     650     660
cctccacaga aacaaggcaa gaaggaaaaa aactttcaca tgtagaatta taaatggaaa
 670     680     690     700     710     720
aataaatttt ctagttttct taaagaccct ggtttccggt ataaagaaat gtcccagctt
 730     740     750     760     770     780
cttagtcacg ggccaatggg ttgttggtac tgaaaaggat gggaaacagg aatctactat
 790     800     810     820     830     840
ttttgatgct gtaatgattt gttcaggaca tcacgtatac cccaatctgc caacggattc
 850     860     870     880     890     900
ctttcctggg aagtttgga aatatataat aatctaggga cttatatgca aacatcaaga
 910     920     930     940     950     960
gttagaaaca tatctttcta taggtattac ataatgatta ttcttagatt tcaaaaagaaa
 970     980     990    1000    1010    1020
aaaattaagt ttaatgatag gatatagtaa taaatagcct cataagtcct tatgttaaaa
 1030    1040    1050    1060    1070    1080
taatcaagga ctgcaagcca gagatcagac aaacacaagt tcctgtggag aagtgcggga
 1090    1100    1110    1120    1130    1140
cccaaaaata agttctaaca gcacacgggg tctccgagca cagttacatt aaaaaaaagt
 1150    1160    1170    1180    1190    1200
agagtccaac tgccaaatgg tttaaagaaa gacacgttta cttatgttat ttataggaga
 1210    1220    1230    1240    1250    1260
ctcctagggt tctaatttca tcttcantcc acaatttgca aataaacttt agaaatctca
 1270    1280    1290    1300    1310    1320

```

Figure 10 (suite)

39/53

fragment3.FMOx

```

gtgatttgtg tgtgggtaca cacatgggtg tgtgtatagc agcatacttc attaccatcc
1330      1340      1350      1360      1370      1380

gaaagtggca aacctcaaat aaatacaata tcatggagg cttccttcca ttttctcttc
1390      1400      1410      1420      1430      1440

cttccttgcc acagggaacac aatctactca aagatattag agtttccatg tctagggtatg
1450      1460      1470      1480      1490      1500

atgtccatag gccgaggaaa attaaagagt gaagggttcag gaggaatata agattaaaac
1510      1520      1530      1540      1550      1560

tcttaatggtt aacgggcagc atattttaatg tttatgagca tgggatcaga acacctggcc
1570      1580      1590      1600      1610      1620

tcaacttact attccactag ttccttacca cttaacttct ttgtctcaat ttcctcttct
1630      1640      1650      1660      1670      1680

tttaaaatag ggacaatagc ccacatgca gggatgttat caagattaaa tagttaaaac
1690      1700      1710      1720      1730      1740

gtgtaaagca tttatcagag gatctagccc acagagttaa cttaataaat attaacatt
1750      1760      1770      1780      1790      1800

attattatcg taacatacat tctcatgcct taagattttt taaggaaact aaagtaagtt
1810      1820      1830      1840      1850      1860

ttagggggct taatgtcaaa aaatgctaaa tggataaatg cacttcaact aggggaatttt
1870      1880      1890      1900      1910      1920

ttaattacaa ctgataatag gtttaaaaag acacaaaagaa aacatcttca taattttctga
1930      1940      1950      1960      1970      1980

aaatcagttc aaacaacttg ccatgttcca cttaggcctg gaccagtttc gaggcaacta
1990      2000      2010      2020      2030      2040

cctccatagc cgggattata agaatccaga agccttcaag gggaagaggg tccctcgtgat
2050      2060      2070      2080      2090      2100

tggtctgggg aattcgggat ctgacattgc tgttgagctc agccgtcttg ctacacaggt
2110      2120      2130      2140      2150      2160

acatgacgta aagggttttg gaaataaacc taaggtaggg ctgtgctact aaatcagtag
2170      2180      2190      2200      2210      2220

ccaaggcaca gaggatggta cttctatgtc acaccacaag agatccacct cttctatgtg
2230      2240      2250      2260      2270      2280

gcccttcaaa tcaaggagga cttgagacat cctccatgtg aagccaggtg atgtggcccg
2290      2300      2310      2320      2330      2340

tgctagtaag gaagtacatt ccactgaatc cagaagtaag tgcattgagtg cgtgtatgta
2350      2360      2370      2380      2390      2400

cagatgagtg tgtatgtgtg tatttcttgt tttcatttta tattctgata acctccaaat
2410      2420      2430      2440      2450      2460

agactagttc ctggtcaggc ttaattcttta tttatttaac agtatttatt ataacgtatc
2470      2480      2490      2500      2510      2520

atgcaaaaaa cactgtgttt accactctga agttctgaaa gatatgcatg acttggtatt
2530      2540      2550      2560      2570      2580

tactaacatt aattcaatca acagcagatg ctcaacaaat attgggcact tactatgctt
2590      2600      2610      2620      2630      2640

actatgtgtc agaactatga taaactaaaa ataaatgcat aaataagtta gactagttcc
2650      2660      2670      2680      2690      2700

tgacttcaag aaagagtcaa tggatggaga tggagtggac aggtacacac agactatcac
2710      2720      2730      2740      2750      2760

```

Figure 10 (Suite)

fragment3.FMOx

```

cagaggagat ggtgagtctt ccagtagaat taggtgtggc aatagcaaca cagggaaaaag
2770      2780      2790      2800      2810      2820
agaatctaac tttagcctgga tgagggtcaag gaagacttcc cagaggactc caagctaaat
2830      2840      2850      2860      2870      2880
catgtatcat cgatagaccc taaagaaaca acatatTTTTT aagaaaacag gttctcaata
2890      2900      2910      2920      2930      2940
aataaattct taaatggatg taaataaaac cttaattttt taaactaaaa attcccttca
2950      2960      2970      2980      2990      3000
gttatcacia agttaaatgc tattttgcaa agacggtaaa atagataagc agccagactc
3010      3020      3030      3040      3050      3060
atctcagggc tgaggcgggt gccatgggtt ggggttgctca ggagaagtcc ttgggggtatg
3070      3080      3090      3100      3110      3120
tgtatagggg gaactggaaa aggcaaccag agacagagaa cagaattaaa tccctgacat
3130      3140      3150      3160      3170      3180
ctcgtcagcc taatttcagc tagagattta gctacacttt tcccacacct agtccactat
3190      3200      3210      3220      3230      3240
caccagccac aaccactggg gctcactgga tcactctggc cctaccagac ttgccatctt
3250      3260      3270      3280      3290      3300
agtctatgag tatgtgaaga ttaaacctc acagttgaac acagagccct gttgttccct
3310      3320      3330      3340      3350      3360
gagggtatgat tctaattcct tcaacaacta cacaccagcc ctccaggggca gtgaaagaat
3370      3380      3390      3400      3410      3420
cctgtctctc ctagttttaa ttttagactt taaaaaaaat tttttttatt ttaagtcttg
3430      3440      3450      3460      3470      3480
ggatacatgt acagaacatg cataggtctg cacatgccat ggtgggtttgc tgcacctatc
3490      3500      3510      3520      3530      3540
aacccttcat ctagatttta agccccacat gcattaggta ttgtctctaa tgcctccctt
3550      3560      3570      3580      3590      3600
cccttagccc tccatccccc cgacaggcct tgggtgtgtgt tgttccctt cctgtgtcca
3610      3620      3630      3640      3650      3660
tgtgttctca tgattcaact cctgnnnnnn nnncttatga gtgagaacat gcagtgttcg
3670      3680      3690      3700      3710      3720
gttttctgtt cctgtgttag tttgtgagg atgatgggtt ccagcttcat ccatgtccct
3730      3740      3750      3760      3770      3780
gcaaaggaca tgaactcatt cttttttatg gctgctagac aacttattta gactcgcctt
3790      3800      3810      3820      3830      3840
ttaaagtgtt tcctacttgg atattgagga aaatgcacgg aagtgcccaa agaagtgtgt
3850      3860      3870      3880      3890      3900
tgtgtttgct tatttcttac agagtaatgc tgaaatctgt gttgtctttc cccaccaggt
3910      3920      3930      3940      3950      3960
cattatcagt accagaagtg cttcctgggt catgagtcgg gtctgggatg atggctatcc
3970      3980      3990      4000      4010      4020
ttgggatatg atgtatgta cccgctttgc atcctttctc cggaaatgtcc ttccttcatt
4030      4040      4050      4060      4070      4080

```

Figure 10 (Suite)

fragment3.FMOx

```

catctctgac tggttatatg tccagaagat gaacacgtgg ttttaagcatg agaactatgg
4090      4100      4110      4120      4130      4140

cctgatgcct ttaaattgggt acttaaaaat ggaaatttnt ttttattcaa aaaagggggg
4150      4160      4170      4180      4190      4200

cactcattta atgaatttat tctctctaga acttactttt gttgtctcat tgagcctaga
4210      4220      4230      4240      4250      4260

aacattaaac tcaaggtttc acagggtgacg gaatatgccc agagaccacg tatggcctgg
4270      4280      4290      4300      4310      4320

aaaacttatt gaaattagtc cagtacagaa aggggtatgga aaaatctgaa atggagatga
4330      4340      4350      4360      4370      4380

cgcaggcaga taaatcaccc tgacatgcat gatgcatttg tgggtggctac aagctatagc
4390      4400      4410      4420      4430      4440

atagaacttt gaggactgaa caaactcaaa ttgggttttg gaagaatatc ttgtccgtgc
4450      4460      4470      4480      4490      4500

ttatgggtgt atgaagacat caataataat acttgcttct caagatgggt gtggtattca
4510      4520      4530      4540      4550      4560

ataatataaa aatataaaaa ttgctttcta aatgataaag ctttaaaaaa attggttctt
4570      4580      4590      4600      4610      4620

cttagtctca atttttctaa tgtgcttcaa aggagcaaat aacaaaatag tgттаatcaa
4630      4640      4650      4660      4670      4680

catgtctcag caagtaggaa gtctcaaaac aaaagtgcac acttcctcca cccctgaaat
4690      4700      4710      4720      4730      4740

gttgacattt ttgcagaacc atcaggaggc atggaacaca taaagtaatg gagagtcaca
4750      4760      4770      4780      4790      4800

actaacgtgg cctgtaagat tagtcagatt catttattta cttctttata gagacagggc
4810      4820      4830      4840      4850      4860

ccaacattta ctaattagga agtcattcca ggtagaagaa tcagcatatc aatagaaaaa
4870      4880      4890      4900      4910      4920

aagaatattt aagttggtaa gaaaagaaag aattgagaaa ttttatctcc tggccccatgc
4930      4940      4950      4960      4970      4980

tagccaaaaa gtttcattgt gtttagagaa agatggtaag aaaaaggagg aactgtaaat
4990      5000      5010      5020      5030      5040

caaaagagca aatgccagat ttaggagcta aactgtcagt ccaaagcact tatactacca
5050      5060      5070      5080      5090      5100

agtcttgcag gctgctataa ccttttaaaa tatgttgatt ttatgcattt aaaatttatgt
5110      5120      5130      5140      5150      5160

ttaacactgt ggcttgcttg acagtagagg gtgggaggag gaagagaatc agaaaaaaat
5170      5180      5190      5200      5210      5220

acttatcagg tactatgctt attaccagc cgacaaaatt atctatacac aaaacccctg
5230      5240      5250      5260      5270      5280

tgacacacaa tttacttata taacaaacca tggaccccca aacctaaaaa aaaagttttt
5290      5300      5310      5320      5330      5340

aaaaattatg tttaatatag taagtcccat agcttgagct ggttaagatt ttttatcttg
5350      5360      5370      5380      5390      5400

taagagtaac tataaattat attttggcct tgccatttag acaattaaaa catagtttta
5410      5420      5430      5440      5450      5460

```

Figure 10 (suite)

fragment3.FMOx

```

gaaattcatt cattctgaaa actaagcttc cttttggaaa gggttccaat taccctaagt
5470      5480      5490      5500      5510      5520
ttctggaggg agaaaggggg aggaaaaaca ggtttcattg tggctctatgt tttgctacct
5530      5540      5550      5560      5570      5580
tgtaaggtaa aagaagaggt tgcaggatta gataaacaga aaatgatgtg gaagtataga
5590      5600      5610      5620      5630      5640
gacaagtttc aggatttaca aggtttcttt gtgtctgaga tacttgagg aaattccgga
5650      5660      5670      5680      5690      5700
atctcaaagg aaacttaaat caaatgaaa tatattgtcc tgaaaaatat tattcctaga
5710      5720      5730      5740      5750      5760
attttgcaa ctaaatgca atatcaaagt tgttacactt ttttgggac acagctgatg
5770      5780      5790      5800      5810      5820
aaagaaaacc aaacatggca ataaaacttc ccaccactgc aagtctgatt tctccatgta
5830      5840      5850      5860      5870      5880
agacaagacg ttaaggttat gataatagt cacttataac aacagtgcct gcatgtgcc
5890      5900      5910      5920      5930      5940
ggaactgttt taagtgtttt aaggataatt gatcatttaa ttttcacaac aacctatgag
5950      5960      5970      5980      5990      6000
gaagattcca tcatcatccc cattttacac ataaagaaac aaatacagaa aagtaacaac
6010      6020      6030      6040      6050      6060
tagtaagaga tggagctagg ttatgaacct gggccatctg cttccagagt tggcgttttt
6070      6080      6090      6100      6110      6120
aaccacttta gtatgtctat aaattagttt tagtctcatt taggaaagga attgccatga
6130      6140      6150      6160      6170      6180
gagaagagag tcagtggcac tcatgtctgat gtttaagtgc ttgatgttat ttcaatgtta
6190      6200      6210      6220      6230      6240
tgggctgttg caggtatttc ttggaaatga gctatttaca gcaagggtgt ttgcctctca
6250      6260      6270      6280      6290      6300
ttgctgtagt tccctgagaa aagagcctgt gttcaatgat gagctcccat cccgcatect
6310      6320      6330      6340      6350      6360
gtgtggcact ctgtccatca agcccagtgt gaaggagttc acggaaacct cagctgtgtt
6370      6380      6390      6400      6410      6420
tgaggatggg accatgtttg aggcattatc ctctgtcctc tttgcaacag gctatgatta
6430      6440      6450      6460      6470      6480
ttcctacccc ttccttgatg agaccatcat gaaaagcaga aacaatgagg ttaccttgtt
6490      6500      6510      6520      6530      6540
taaaggcatc tccccccac taatggagaa gccaaccttg gctgtgattg gcttggttca
6550      6560      6570      6580      6590      6600
gtcccttggg gctgccatcc ccacagcaga cctgcaagcc tgggtgggctg cttaaagtatt
6610      6620      6630      6640      6650      6660
tgcaagtagg tgggccattc tgtctttcat tcattttatc aatgaacatt tactgaacac
6670      6680      6690      6700      6710      6720
ctgcttatatg caaagcactg tgctagggat acaatgagaa caagacaaac atgttccttg
6730      6740      6750      6760      6770      6780

```

Figure 10 (suite)

fragment3.FMOx

```

acctctcaag gcttaaaatg ggggtgtggg gatgccataa taggggaaat ttgggggggt
6790      6800      6810      6820      6830      6840

tctagtgaag ggagttggac tgttgacacag agcaaacagt atacaggaag tcataaaggt
6850      6860      6870      6880      6890      6900

gagggaaagc atgaaatgtg taaggaccca gaaacatttt ggtggaagg aatataaagc
6910      6920      6930      6940      6950      6960

agaggcaggg agtggaaga aatataggtt tataagccac gttaaagagc ttaaacttct
6970      6980      6990      7000      7010      7020

catagggatt aaggacttcg caagatttta agcaagaaaa aaatagcaga ggataactgc
7030      7040      7050      7060      7070      7080

aatgtcaggc tacattataa agattggaag ggccctgggt agggttggag gtgtgccaga
7090      7100      7110      7120      7130      7140

aacctcactg gtgtcaactt ctgtcagaat aacaaagtca ggccactctg attctcatga
7150      7160      7170      7180      7190      7200

caatcttttt cttctctccc tctactctag acctcatggt ctccaggggc tacaagtatg
7210      7220      7230      7240      7250      7260

cttatgtgag gaaatcaaga atatgaggat tacatggaga aaggcaatgt ctcaaataa
7270      7280      7290      7300      7310      7320

ttaatttact ccagtcatac tgaatattat cattattatt gaaaagtgtt cttttattca
7330      7340      7350      7360      7370      7380

ggtatttctc aaaatattga ccaatatagg tataacttac ctaacataac taatccataa
7390      7400      7410      7420      7430      7440

aaacttacac tatttgtaat taacaaacca ttacaatcat ggaatatatg tatatatatt
7450      7460      7470      7480      7490      7500

gtctaaaact ttgtagataa ataaatttct atttcaaata caccatgaaa gatcatcatt
7510      7520      7530      7540      7550      7560

taataaaacc ccatcatgaa atcttttgta aagggtgctc ctgcaaaata cttctattgc
7570      7580      7590      7600      7610      7620

ctttttctct cgaaaggcac aacaatgcc aagcctggg gtattatgag aagactggat
7630      7640      7650      7660      7670      7680

atagttcata aacctaagaa atttacatga agcaaagtgt atcatttatt tattcagcaa
7690      7700      7710      7720      7730      7740

atacttactg aacacctact atgtgtcagg ctctaacctg gcacttagga cacaacaaca
7750      7760      7770      7780      7790      7800

aacgaagcag aacaaaattc tggcctctta ctttctagca ggggtgtccag ccaatatcaa
7810      7820      7830      7840      7850      7860

tcatagggta ctaccaggtt gacataagac actaacgatg actgggaaat attcatgcac
7870      7880      7890      7900      7910      7920

tgcaaatttt agagtaactt tcttccactg ttacaaaggc aaataagcta ccatcaccag
7930      7940      7950      7960      7970      7980

ttaaaagaag ttgcattgat gtagtgaaat tcacaaaaag ctaaaacttg tctgtgtccc
7990      8000      8010      8020      8030      8040

cttaaaacac cttgcatagt tgcagaagat gtttaaaatc ctatgcttcc ttccattacc
8050      8060      8070      8080      8090      8100

```

Figure 10 (suite)

fragment3.FMOx

```

tcatttaaaa tggcagaaac cttaaaggga actgtttttac cagattcttt cttcagagaa
8110      8120      8130      8140      8150      8160

gttttaggaa aaggatacag aaaaaaaagg aagaaattat taagctatta tatgcatgaa
8170      8180      8190      8200      8210      8220

gtgtactgag cacatatgtt gaggattagg tcctctataa tgttaccgaa ataagagact
8230      8240      8250      8260      8270      8280

gagtgtattg aagctacaaa tgtctctgct gtcactatct cactacaggc cagctttgcc
8290      8300      8310      8320      8330      8340

atccccaaa gggttcattaa cttttcagat ctttggtttct atgaactggg attttgctaa
8350      8360      8370      8380      8390      8400

agatatcaaa gacatctcca gctcctctta atacaaaagt tttcaggaat acagtttata
8410      8420      8430      8440      8450      8460

aaaaccaaag gatttccatc atatgtcatt atatatctct gatttgtggt tttcaatatt
8470      8480      8490      8500      8510      8520

tttctcttca tttctttttct agactcatgt accctgccaa ccacgaatga aatgatggat
8530      8540      8550      8560      8570      8580

gacactgatg agaaaaatgg gaaaaaactc aagtggtaag cagctaactg tacttgctaa
8590      8600      8610      8620      8630      8640

tagagcaagt tcctaaaatg tgcctttatg tgtagaaaaa cattaatatg ctttaatat
8650      8660      8670      8680      8690      8700

gtcatttagtc agagtttaca ttttctgaac acttgcaata atcaaaaaat gtttagatag
8710      8720      8730      8740      8750      8760

taaacagtc tcaacttctt cttgtgtaac tcaagaatag aggttttcta tcagggataa
8770      8780      8790      8800      8810      8820

ttttgccctc caggtgacat atggcaaaat ctggagacac ttttggtcat tgtgagtggg
8830      8840      8850      8860      8870      8880

gagggcatgc tatcagcatc ttatgagtag agaacaggga ttctgctaac catccaacaa
8890      8900      8910      8920      8930      8940

tgcagagcac agttcaccaa aacaattatc tggctcaaaa tgtcaatagt gctgagggtta
8950      8960      8970      8980      8990      9000

agaaacaact ctataaatga ctacagttag cttttgaaca acacagggtt gaattatatg
9010      9020      9030      9040      9050      9060

ggtccactta tacatggatn ttttcaatta acataantcc agattgggca tgggtggctca
9070      9080      9090      9100      9110      9120

cgccctgtaat cccagcactt tgtgaggctg agacggggcg anaaactgag gtcaggagtt
9130      9140      9150      9160      9170      9180

cgagaccagt ctggccaaca tggtaaaacc ctgtctctac taaaaatata aaaaaaatta
9190      9200      9210      9220      9230      9240

gtcgagtgtg gtggtgtgca cctgtaatcc cagctactcg ggaggctgag gcaggggaat
9250      9260      9270      9280      9290      9300

tgcttgaatc agggaggtgg aggttgcaqt gagccaagat cgcgccactg cactctagcc
9310      9320      9330      9340      9350      9360

taggtgacag agtgagactc catctcacia aaaaaaaaaa aaaaaa
9370      9380      9390      9400

```

Figure 10 (suite)

fragment4.FMOx

```

aaaaaaaaa aaaaaaaaaaag ccacaaaaggc cattttattaa taaggagcag aagtgaacac
  10          20          30          40          50          60
caagattagg caggaannaa taagctaact actgttttgt gcaaatgcag tcagatttat
  70          80          90         100         110         120
gatcaggatg gcccttacct atgaagctac cccctcaaac cttgaaggga aaagatgaat
 130         140         150         160         170         180
atcagcttcc tatcttttgg ttatacaaga ccccttttct ggattagctc tgtcaatgct
 190         200         210         220         230         240
ttgtccctga agtcagaaaag tccttgccaa taagagactg ccttttaaag tttttttgat
 250         260         270         280         290         300
atagacaatg cccctgacca ccagaaaccc catgagttca acatggaagg catcgaagta
 310         320         330         340         350         360
gtctaatttc ccccaaacac aacatttctaa ttcagccttt atatcaggga gtcataagga
 370         380         390         400         410         420
cctttaaggc tcatcacata ccatactcta tggaaaagat agtcaatgct gtggaagata
 430         440         450         460         470         480
acccaacaga gagaacatca tgaaagtctg gaaggattat accattgaag atgccctaatt
 490         500         510         520         530         540
tgttatagaa aaagccatga aagccatcaa tcctaaaaca acatatttct cctggagaaa
 550         560         570         580         590         600
actatgtcca gatgttatat atgacttcag aggatttaca acagaccagt cacagaaatc
 610         620         630         640         650         660
atgaaaaaga ttatggatat ggcaaaataa aaagggtgagg gtgaagggtt tcaagatatg
 670         680         690         700         710         720
gatcatggag aaattcaaca gctaatagac accactaata gacactttta attccacact
 730         740         750         760         770         780
agaggaacta aaagatgact tgatggagat gagtccctcc aaagcagtgc cagatgagaa
 790         800         810         820         830         840
cgaagacata gaaaaagcca tgccagaaat aaattgacat tagatcatct ggcagacagg
 850         860         870         880         890         900
ttccagttat ttaagacttc ttttgacttc ttttatataa catggaccct tctatgatac
 910         920         930         940         950         960
aggcactgaa actaaagcaa atgatatagg aaggattact actatataga aaatttttag
 970         980         990        1000        1010        1020
agaaataaaa aagcaaagtc agacagaaat tataatatat ttccataatt acaccaatgg
 1030        1040        1050        1060        1070        1080
gcctgcctct cctgccccca attctacctc ctccatctct tccgcttctg ccaggccctga
 1090        1100        1110        1120        1130        1140
aacagcaaga ccaaccctct ctgtttctcc tcctactcct cagccctactc aacataaaga
 1150        1160        1170        1180        1190        1200
tgataaggat gaagacattt atgataaacc acttccactt aatgaatagt aaacatatct
 1210        1220        1230        1240        1250        1260
tttcttctct ataattttct taataacatt ttcttttctc catattactt tattgtaaga
 1270        1280        1290        1300        1310        1320
atagtattta atacgtatga catataaaat atgtgttaat caactgtttg cattattggt
 1330        1340        1350        1360        1370        1380
aaggcttcca gtcaacaggt tattaagagt taagtatttg gggagtcaaa agttatacat
 1390        1400        1410        1420        1430        1440
ggatttttga ctgcaagagg gctcaatgcc cctaaccctc cagttcttcc aaagtcaact
 1450        1460        1470        1480        1490        1500

```

Figure 10 (Suite)

46/53

fragment 4.FMOx

```

gatataaggaa gtttctttac tttttcaagc atttaacatt gcattgatat gtcaacctaa
1510      1520      1530      1540      1550      1560

aggaaaaacac tgaggcaaat ttaatatataa aagagaattg atttggggcca agtttgagga
1570      1580      1590      1600      1610      1620

ctgcaaccca ggagcacaga gtcaaattgc cctgaatatg cactccgttg gcagcagtta
1630      1640      1650      1660      1670      1680

caagtagggtt ttttaaagga aatacaaaaag agtcaacttc taagttgttt accaagaact
1690      1700      1710      1720      1730      1740

tacattaaaaa ttatataagc tattgattgg ctatatactg ttcttcgtat cacaaattct
1750      1760      1770      1780      1790      1800

acgaacatga agatgatgag tgagacagct agtcaggaat aaaaatgcct ttaacaatt
1810      1820      1830      1840      1850      1860

gccgccaggc ttggtagagg gcagcatgac aagtcaccata cacgtggctc tctcagcttg
1870      1880      1890      1900      1910      1920

ataaatttttg catacctcac atagtgcata ctactctgag ctatttttct tctctcacat
1930      1940      1950      1960      1970      1980

tgaatgccac aatgtagtca cccattcagg gcctagagaa gaaaagaaat ggaaccctca
1990      2000      2010      2020      2030      2040

gattcaacaa aacctctcct gcacaacttc agccagttga cgaacaactt gcagagtttg
2050      2060      2070      2080      2090      2100

gcacttttat gtgctaacaa ttcattgcagc ttgataccct ttctttttaga gccagtaga
2110      2120      2130      2140      2150      2160

aataaaaaatg aggaaataga gaggttataa tgttcattctt attgctttaa tgataagctg
2170      2180      2190      2200      2210      2220

ctcttcagag tttcaaaaag caaattacac catattccaa ctaaaagaac tatagaggcg
2230      2240      2250      2260      2270      2280

gaaaggaggt gatctctttt ctctctgtca taaaaggtaa tggccaacac ccctataaca
2290      2300      2310      2320      2330      2340

aaagacaggt taacaagaga aaacgtgaca gatttattac gtgcacatgt gtgcatgaga
2350      2360      2370      2380      2390      2400

gccttacaaa acatgaactc aaaggagggc cagatcattc atgttttaat attctcttca
2410      2420      2430      2440      2450      2460

ctgggggttag gggagatgga agtgtaaaag taaatgattt ttcagaggaa attaataagt
2470      2480      2490      2500      2510      2520

ccaaagaaca cagattagac caagtttctc tgggcttttg gggagggtgta atcacccaac
2530      2540      2550      2560      2570      2580

agattcatct tgctcactgc ccagaaaagc tgatgccctg agaacagcag gtttttccaa
2590      2600      2610      2620      2630      2640

tagagagagt ttaataaaca cacagctgtc agaggcattt gaaccagagt gactccatct
2650      2660      2670      2680      2690      2700

tgaatagggg ctgggtaaaa tgaggctgag acccactagg ctgcattccc aggaggttag
2710      2720      2730      2740      2750      2760

gcattcttag tcacaggatg agaaaagagg ccagcacaag attcagggtca caaagacctg
2770      2780      2790      2800      2810      2820

gctgataaaa caggattcag taacgaagct ggccaaaacc caccaaaaacc aagatgatga
2830      2840      2850      2860      2870      2880

aaaaagtgac ctctggtcgt cctcactact cattatatgc tgattataat gaattagcat
2890      2900      2910      2920      2930      2940

gctagaagac actcccacca gcactgtgat agtttataaa tgccatggca atatcaggaa
2950      2960      2970      2980      2990      3000

```

Figure 10 (suite)

47/53

fragment4.FMOx

```

gttaccctat atagtctaaa aaggggagga accctcagtt ctgggaactg cccattttctt
 3010      3020      3030      3040      3050      3060

tcctgtaaaa cttatgaata atccaccctt tgtttaacat gtaatcaaga agtaactata
 3070      3080      3090      3100      3110      3120

agtatactca gttgagcagc ccatgccact gctctgccta tgaagtagcc attctttttat
 3130      3140      3150      3160      3170      3180

tcctttactt tcttaataaa ctgcgtttca ctttatggac tggccctgaa ttctttcttg
 3190      3200      3210      3220      3230      3240

tgagaggtcc aagaaccctc ttttggggtc tggccagcta aacggaagga caggagttta
 3250      3260      3270      3280      3290      3300

ttactactca aatcagcctc catgaaaatt cagaggctag atttttttaa ggatagtttg
 3310      3320      3330      3340      3350      3360

gtagtcaggg gctaggggat ggggaatgct gattggttgg gtcagggatg aaaccatagg
 3370      3380      3390      3400      3410      3420

gagtcaaaagc ttgtcttctg gtcttctctg gaggagacca catgacaaga tgaaccagtt
 3430      3440      3450      3460      3470      3480

taccagtctg ggtagtgcga gccggcccat cagaatgcag ggtctgaaaa atatcttgag
 3490      3500      3510      3520      3530      3540

caccaatggt aggttttata atggtgatgt tatccatagg agcaattggg gacttctgac
 3550      3560      3570      3580      3590      3600

tgcatgactc ctgagcccta atttcttctc ttgtggctaa tttgttagtt ctacaaaagc
 3610      3620      3630      3640      3650      3660

agtctgatct ccaagcaagg aggggggttg ttttgggaaa gggctgttac catctttgtt
 3670      3680      3690      3700      3710      3720

tcaaagttaa actgtaaact aaatgtctcc catagttagc ttggcctatg ctcaggaatg
 3730      3740      3750      3760      3770      3780

aataatggca gcttgagat tagaagaaag atggagtaat tacatttttt ttccacattt
 3790      3800      3810      3820      3830      3840

ttttcactgt cacaattttt ttaaagggtg tttcagaggt aacatcacag gacatgggag
 3850      3860      3870      3880      3890      3900

actaaagggg gaaaagtatg tcaaacaaag gctgtcctgt tctgcagacg aaacctcaca
 3910      3920      3930      3940      3950      3960

gaaagcaact ctcagagtca gtgcctatg atgaaagttt ctctgtcaga cattcagcag
 3970      3980      3990      4000      4010      4020

tgcttgactc tcagtctctc tctcctgcaa gttaattctt cctagagtgg gcaagggagg
 4030      4040      4050      4060      4070      4080

cctccgagaa agcctagttt ccattctctg tttacttcct tttattttct ccacagataa
 4090      4100      4110      4120      4130      4140

aaatctcctt cacaaaaggc agcttttcag ggctgtttct gtctgcaggc cctctgaata
 4150      4160      4170      4180      4190      4200

gccatctcaa aatctgtcaa cgaagtgtat attttgcagt aaaatatatt ttgttttctt
 4210      4220      4230      4240      4250      4260

tagtatgaaa caatttatat tattagatta caggagtatt aaaaccatcc atgatctcac
 4270      4280      4290      4300      4310      4320

ttttaaacaa accaatctga aagtctaaca ttggggcaga ttctaagcaa tgtcttataa
 4330      4340      4350      4360      4370      4380

agaataatta tgtgttaatg agtaaaacta gtttaattagt ctccttaaac cagagggtca
 4390      4400      4410      4420      4430      4440

gtttactcca ggccacatgg tcaaaggcaa aagtccaaca ttacatcaaa ctcaaataga
 4450      4460      4470      4480      4490      4500

```

Figure 10 (Suite)

48/53

fragment4.FMOx

```

gattaggaag gaggagaaaa gcagctcact tagctaaaga aaaaacaata aattcaattt
4510      4520      4530      4540      4550      4560

tgtggaaaag gagggcataa atggaggtgc tatctaaaat gttatttttc tgaaagaaaa
4570      4580      4590      4600      4610      4620

aataagaaat taatgtctct atttgcaact gtaacactta ttccagtatg ttctcttctt
4630      4640      4650      4660      4670      4680

tcttcatgtt tggccagagc cagactttgc agacagatta catcacatat gtggatgagc
4690      4700      4710      4720      4730      4740

tggtgtcttt catagggggc aagcctaaca taccatggct ctccctgaca gatccccgcc
4750      4760      4770      4780      4790      4800

tggtccctgga ggtgtacttt ggtcccttgc gcccatacca gtttcgactg atgggaccag
4810      4820      4830      4840      4850      4860

ggaagtggga tggggccaga aatgccatcc tgaccagtg gaaccggaca gtgaagccaa
4870      4880      4890      4900      4910      4920

ccaggacaag agttgtcagt gaagttcagc gaccccatcc cttttacaat ttgcttaaaa
4930      4940      4950      4960      4970      4980

tgctttcatt cccattactc cttctggctg ttacacttac attttattaa tgagaaagtc
4990      5000      5010      5020      5030      5040

tttgaggtct caaaattcag catagaagtg taatcacaca atacaacaca caccacacat
5050      5060      5070      5080      5090      5100

acacacacac aatcacaca tagttcctct ctcctttcct gaagatatga aaatcagtct
5110      5120      5130      5140      5150      5160

tggtccattt gaattaaagt ataatgaaaa tggaaaatac tcagccctctc tctctctgtt
5170      5180      5190      5200      5210      5220

gggaatctgt tctctaaaag gcttttcaca tgctgaattg gcaaatttgg ggatgcttaa
5230      5240      5250      5260      5270      5280

gataagacag gaagttgaat aagcatgagc acagattttc aggtctgaaca aagtttaaca
5290      5300      5310      5320      5330      5340

aaatgtaaaq gcattttatt ttgtccatta tagtatatta accaatgtat ttttcatatt
5350      5360      5370      5380      5390      5400

aactgaagtt attgacatta aatcaacata gagagataag ttatatacaa tacaacaacg
5410      5420      5430      5440      5450      5460

aagataaaat gaacacattg cttgaaagat acatactgcc aaagctctcc caagaagaag
5470      5480      5490      5500      5510      5520

catataacct gaatagatct aaatctatta aagtacttga atttgtagtt taaaaactct
5530      5540      5550      5560      5570      5580

tccacaaaag acaggccaag aaccaggtct tcaactgtta attctgtcaa acatttgaga
5590      5600      5610      5620      5630      5640

aagcaatata ccaactatat gcaaaatttt ccagaagatt aaagaggaca gaataatcct
5650      5660      5670      5680      5690      5700

cagctcatca tagccctgac acctaacaac gacaaacaca ttctaaaaaa ataaaatata
5710      5720      5730      5740      5750      5760

gatcaatatc cttcatgaag atatatgcaa aaattcttaa caaatgttaa tataatcaaa
5770      5780      5790      5800      5810      5820

cagtttataa aatgaatagc ataaccaagt aagaattatt ctggaaaagc aatgttggtt
5830      5840      5850      5860      5870      5880

taatagtcaa tgtaattcac cacatttgta aactaaaaat gaaaaaccat aggatcatct
5890      5900      5910      5920      5930      5940

caatagatgc ataaaaagca tctgacaaaa gcccaacatcc attcctgaag aaaactctca
5950      5960      5970      5980      5990      6000

```

Figure 40 (suite)

fragment4.FMOx

```

acaagtgat agaaggaac ttcctcaaca tgaaaaaggg catctatgaa aaacctacag
6010      6020      6030      6040      6050      6060

ctaattatcat actgaatggg ggaaaactaa acgctgtatt tctaagagca ggaacaatat
6070      6080      6090      6100      6110      6120

aaaggccttg gttctcacca cttctagtca acatcctact ggggtttcta gccagtgcga
6130      6140      6150      6160      6170      6180

taaggaaaaa gaaagaggaa taaaagatat caaaagtggg atggggagaag tagaactttc
6190      6200      6210      6220      6230      6240

tttatttctta gacaacataa ttgtctatac agaaattatt ttgaaatcta cataaaaatt
6250      6260      6270      6280      6290      6300

acagaacaaa taattatatt tctatatact aacagttaac agtagaaaat tagaaaattg
6310      6320      6330      6340      6350      6360

taaactatta ttataacat ccaaaaatat gaaatattta ggataagcca gcataagaaa
6370      6380      6390      6400      6410      6420

ggtaaaaacc atacactgaa agctaaaaaa catagctgag agtaattaaa ggtctaaata
6430      6440      6450      6460      6470      6480

aatggagaga tacactatat cctgggttga accaccgata ttgtgggtgat gtcaattatc
6490      6500      6510      6520      6530      6540

ctcaagttga tccataatth caatgcaata ctaatcaacc ccctaacatg cttttttata
6550      6560      6570      6580      6590      6600

gaaattgaaa actgatttcta aaatgtataa ggtattttaa ggacctagaa ttgttaaaaa
6610      6620      6630      6640      6650      6660

ataactttga aaaggccggg cacaaaggct cacgccagta atcccagtat ttcgggaagc
6670      6680      6690      6700      6710      6720

tgaggcaggg ggatcacctg aggtcaggag ttcgagacca gcctggccaa catggagaaa
6730      6740      6750      6760      6770      6780

ccccatctct actaaaaata caaaattagc cagacgtggg ggtgcatgcc tgtaatccca
6790      6800      6810      6820      6830      6840

gctactcggg aggtcgagac agggagaatcg cttgaggcgg aggttgacgt gagccgagat
6850      6860      6870      6880      6890      6900

tgtgccattg cactccagtc tgggcaacaa gagccaaact ccatttcaaa agaaaaaaca
6910      6920      6930      6940      6950      6960

aatgtggagc attaacacta cctgacttct gatttcagaa tttcttattt tatgaagcta
6970      6980      6990      7000      7010      7020

cagtaatcaa gtcagtatgt tatgggtgtga agataaacia taggtcaata aaaaagaatg
7030      7040      7050      7060      7070      7080

gaatgtctgg aaatagaacc acacaatatg gacaaagggtg aaaattcagt aaagacagga
7090      7100      7110      7120      7130      7140

gagtcctttc aacaaatgat gttggaacca ttggatatca tgcaatatgg aaaaaaaaaa
7150      7160      7170      7180      7190      7200

ccttcagtca atactttgta ctatatacaa aaagtaactg aaagaggatt atagaactaa
7210      7220      7230      7240      7250      7260

aaccctaaaa cttctagcag aaaatgaaaa agaaaatctt tgggttaggc aaaaatgtct
7270      7280      7290      7300      7310      7320

tagatacaat actaaatgca caatattcat ttttaaaatt gtgtaaattg aacttcacat
7330      7340      7350      7360      7370      7380

aaattaagaa cagttctagc atattctcac tcacagggtg gaattgaaca atgagatcac
7390      7400      7410      7420      7430      7440

atggacacag gaaggggaat atcacactct ggggactgtt gtggggtagg ggtagggtgg
7450      7460      7470      7480      7490      7500

```

Figure 10 (suite)

fragment4.FMOx

```

agggatagca ttgggagata tacctaatagc taaatgacga gttagtgggtg cagcgcacca
7510      7520      7530      7540      7550      7560

gcatgacaca tgtatacata tgtaactaac ctgcataatg tgcacatgta ccctaaaact
7570      7580      7590      7600      7610      7620

taaagtataa taaaaaaaaa gaacagttct aatattcctt taggaaagaa tacagtaagt
7630      7640      7650      7660      7670      7680

acattgatac ttatgacatc cacattatct gtatctaata aaagcaggat gaatattctg
7690      7700      7710      7720      7730      7740

actcaatttg tctaattgga gagatggata agtcatagta gaaaaagttc agacttgaaa
7750      7760      7770      7780      7790      7800

agaaaaatgt ataccaaact atctcaatct gtttttagaa tatgtataat ttatcctagt
7810      7820      7830      7840      7850      7860

aatcaaaatt ggttcaaacc agatgtgtct ccaaagtatt cttgtgtaca tgtgtatgtg
7870      7880      7890      7900      7910      7920

tacacacatg ctagtgttgt gttgtgttta gtatgatgct agattgtgta acagaaagag
7930      7940      7950      7960      7970      7980

tctgtaagaa ttataagctg attttccaaa taccctatgc acatatttct gagtcctctg
7990      8000      8010      8020      8030      8040

ttcaatctgg ggactaccaa tgaagactaa tctggaatct tggagagaga cacttgggaa
8050      8060      8070      8080      8090      8100

gaacaagaaa aataatattt tgagtggggg tctcttttct ggaacatggt tcagggtgtg
8110      8120      8130      8140      8150      8160

agaatctcat tccctcctca gacctagtac tagagggatg ttctcctcaa caaagactgt
8170      8180      8190      8200      8210      8220

agtgaagatt ggctaaatgg actgctgagc ccttcagact tagtttattt gagccaacta
8230      8240      8250      8260      8270      8280

gaaacttttt tgtaattcag acttcttgat attttattca atcatatgga taatattaat
8290      8300      8310      8320      8330      8340

ttgcactaaa ttaaaactga tggaatgttg cttataacct ctgtgtattt ggtcctgtgt
8350      8360      8370      8380      8390      8400

agtaagtaga cataagtgtg gtccttccca tcatctttac caaccatgca aaataaagtg
8410      8420      8430      8440      8450      8460

agaacaaaga gccgttaagg agtgacaaaa ggcgaggagg catgctaact caaagctaac
8470      8480      8490      8500      8510      8520

tcatgcacct cattctactg attgggtagc ctgctcatct gtgaaatcta tgtttgtgtg
8530      8540      8550      8560      8570      8580

tatatacaat ctgtaattga atctagtgtg aggactacac cagacaaatt cattaataaa
8590      8600      8610      8620      8630      8640

ctaactaaat aaaacccttt taccagtga tcccaaaga caatcagctt ttcagtcaga
8650      8660      8670      8680      8690      8700

ttaatagcaa atctcacagt taagaaacta ctatgttttg catcatcaga cacagaagaa
8710      8720      8730      8740      8750      8760

accatatttt taatgtttta ctttgcctct aaaatcctag cataaattag catattttta
8770      8780      8790      8800      8810      8820

aaaaagggtt aagagtagta gtttatgcct ggaatcccag catttttagga ggcccaggca
8830      8840      8850      8860      8870      8880

caaagattgc ttaagcacag gagtttgaga ccagcctaag caacatagca aaacctgttc
8890      8900      8910      8920      8930      8940

ctacaaaaaa attaaaatta aaaaattagc cagggtgctat gactcagatg tgtgtccctt
8950      8960      8970      8980      8990      9000

```

Figure 10 (suite)

51/53

fragment4.FMOx

```

ccaaatctca tgttgaaatg tgatccccag tgttggaagt ggggccttgt gggaagtgat
 9010      9020      9030      9040      9050      9060

tctctcataa atggcttggg gccctcccca gggtaatgag tgagttctca ctctattagt
 9070      9080      9090      9100      9110      9120

tcacgagagc tattatagtt gtttaaaaga gactggcacc tcctctcttc tctcttgctt
 9130      9140      9150      9160      9170      9180

caattctcgt gctccctttc acaccatgtg acatgcctgc tcccccttca ccttctgcca
 9190      9200      9210      9220      9230      9240

tgattgcaag catcttgagg cctcaccaga aacagatgtc ggcaccatgc ttcttgactt
 9250      9260      9270      9280      9290      9300

gtctgcagaa tgatgaacca aaataaacct cttttcttta taaattaccc agtctcaggt
 9310      9320      9330      9340      9350      9360

atctctttat agcaacacaa aatgtgtaac tggcaaaagg gtcctgctgc ttgccacttt
 9370      9380      9390      9400      9410      9420

tagaaaagatg ccaacataac aagagcgaga tgtgacagag aaagagttag ttttatcatc
 9430      9440      9450      9460      9470      9480

tgtgttagca aggaaaagag tgggcagaat tctttcccaa aatttccact cttcaatttg
 9490      9500      9510      9520      9530      9540

tgggaggaac acagggtttt ttttttaaga gacagggtt tggaaatgcag gagaggaaaa
 9550      9560      9570      9580      9590      9600

gggcagctac acagatcaaa aggcaggccc ttagacatgg cacttgcttc tgtatatacc
 9610      9620      9630      9640      9650      9660

agtctcttta aggatataaa ccacaatctg gtttatacat tatgttgatg attataatac
 9670      9680      9690      9700      9710      9720

attgataatt aacaactgta tatctggatg cagtgtcata atatggcttt ttagtaaaaa
 9730      9740      9750      9760      9770      9780

gtacttgaag gtaaaatgaa taattagatt aaaaaatgaa gccacaagga aatgatagct
 9790      9800      9810      9820      9830      9840

gaaattttaag ccacaaggaa atgatagctg aaatttctac gcgtatcttt aaaatttaca
 9850      9860      9870      9880      9890      9900

tcttacattg aactgcagtg atctaggtgc cttcaactgc cctgtcctta gatttaataa
 9910      9920      9930      9940      9950      9960

ggtgtagata gcttcacaga ttaccttttc ttataataaa agaagtataa ccagagcctt
 9970      9980      9990      10000      10010      10020

tatgactaca tttgttaatt tctcctctct gaaataagac aatgtaatct tcagggtattt
 10030      10040      10050      10060      10070      10080

ggtttactaa tatatcacag attcatcacc accccaccca actaatcatc aaatgagttt
 10090      10100      10110      10120      10130      10140

cgcagctaag aactgactca tattctatac ctgccttctt aatgtattaa gaaaatcctt
 10150      10160      10170      10180      10190      10200

aagctccatt aggccttggc caatatgaga catccaaatc cagatacttt tgggactaga
 10210      10220      10230      10240      10250      10260

ctgaaaaaat aagttagtaa agcagataga ttgcatatta atcattctgt tggcacatta
 10270      10280      10290      10300      10310      10320

aacaaagaga aatcatcctt tacttccgca aaaatatattg caggcccagg acaaccagat
 10330      10340      10350      10360      10370      10380

tttctaaact ttttagagaa gccttgattt gatatttcca tctgttattt tctggatttt
 10390      10400      10410      10420      10430      10440

taatgtttag caacaacttt taaaaaaaga tttgtaggtc ttagatgggc tggagaaaac
 10450      10460      10470      10480      10490      10500

```

Figure 10 (suite)

52/53

fragment4.FMOx

tcagtaaggg	cagatttcag	ccgaagaacc	tccactttct	ctcaccttct	ctggaacaaa
10510	10520	10530	10540	10550	10560
tatgtattgc	catcttccag	ccaccattct	cagggttcag	agggagttta	aaggctctct
10570	10580	10590	10600	10610	10620
aatttaaccc	ttggttgata	cttgtgttct	ttctacaata	tccagcctcc	tctttaaacac
10630	10640	10650	10660	10670	10680
ctctttggac	acaaaattca	ttcattaccc	tctggagcag	ccttcagata	actgactgcc
10690	10700	10710	10720	10730	10740
tctggcattt	acccagtaca	gccaccattc	attcacagtg	gcgaacacca	gccctgtgag
10750	10760	10770	10780	10790	10800
atagaaagag	gttcaaattg	gggagtcaca	gctacctgct	caaaaactca	aagcaaaaaca
10810	10820	10830	10840	10850	10860
ctgaactgat	tccaaacagg	agtcttcata	acttattaga	atatttaaaa	aggaataatt
10870	10880	10890	10900	10910	10920
atcttctagga	atctgttata	tggatatattt	tgcattaaag	caaaaagatg	gatgaacaag
10930	10940	10950	10960	10970	10980
gaaaatcatt	gcagtctcgt	ttataatcat	gaaaacctgg	aaacaattca	aatatctata
10990	11000	11010	11020	11030	11040
atagtataaa	ataacataca	gtaaaatata	ggagatcaac	gtatatatac	tgacatggaa
11050	11060	11070	11080	11090	11100
ctcttttcaa	aatatattgt	ttaattgaaaa	aagcaagtaa	caaagttaaac	aatattttctt
11110	11120	11130	11140	11150	11160
catttatgta	aagtctagta	tctgtgtata	taaatgttat	ttttacacag	aaggaaatgt
11170	11180	11190	11200	11210	11220
ctgaaagcca	attgtcatca	gtagttttct	ctgaagatta	gactggggct	ggaaatggga
11230	11240	11250	11260	11270	11280
ggaaggagta	ggaaccgcac	ttttgtttct	acatactttc	gcattgtttg	agttttttca
11290	11300	11310	11320	11330	11340
ttgaagacat	tttcatacat	gacttggttt	atatgtaaga	aaataatttt	ttaaataagg
11350	11360	11370	11380	11390	11400
accatttagt	ctagaaagca	cctacaagat	gcctgattgg	tgaaccttgt	gaagacaaga
11410	11420	11430	11440	11450	11460
gaaaaatgaa	agattttgtg	tatgaatttt	agtctcaact	gcaaaatatt	tttctgcttt
11470	11480	11490	11500	11510	11520
ctcaatgaca	ctacttggtc	ctgtgtgtca	tgttaccact	ccctgtgttc	agatgactgt
11530	11540	11550	11560	11570	11580
gtttgtgcat	tatataacaa	tattaattgt	gctatttgcca	gagagggctg	tgactaatga
11590	11600	11610	11620	11630	11640
ggtatatact	aattagaatg	cactttacca	gaatacacag	ttgtggataa	gccaacaaaa
11650	11660	11670	11680	11690	11700
ggttgagcgc	aggttggtgaa	cttctcctta	cctttcctga	aaacaacatg	aattgtcatc
11710	11720	11730	11740	11750	11760
ttccagtcac	cattctcaaa	gttcaaaggg	agttcaaaga	tctcagccgc	tttgcctgtt
11770	11780	11790	11800	11810	11820
acatgagaaa	gggaagaata	atggaatcgc	aaaagtcccc	aggtaatgca	tgtaattttt
11830	11840	11850	11860	11870	11880
tcataatgca	tcacccagat	caatgctgag	atctggaaaag	agaggaccat	aggttgtgat
11890	11900	11910	11920	11930	11940
aatgtcctgc	cacttcccca	tctatcctca	actctgctgc	aagtctggcc	atagcacgcc
11950	11960	11970	11980	11990	12000

Figure 10 (Suite)

fragment4.FMOx

```

ctgctcaaaa tccttcaaag gcttctcatt accttttggg taatgcccaa gctcaagagc
12010      12020      12030      12040      12050      12060

aaatgaggcc ctctaccatc tggctctact ctttctcatt tttacctctt ccttgctccc
12070      12080      12090      12100      12110      12120

tcttaataact ctggtagtat tggactgggtg atatttccta aactcatacc ataataatcc
12130      12140      12150      12160      12170      12180

tcacctcttt gctccaattc ctgctgttcc cctgggcaca gataccttcc atcacccac
12190      12200      12210      12220      12230      12240

acacatctct accaccatag ccttgggcaa ctctgctctt ttgttattta tctttattct
12250      12260      12270      12280      12290      12300

ttctgcacag tggcttttcc tggggaccat tgtcttcatt cctctttgcc catacctatc
12310      12320      12330      12340      12350      12360

cttacaatata catgttgaaa tgattactac agtcaaggca attaatatat ccatcatctc
12370      12380      12390      12400      12410      12420

acctttccct aattgggtgtg tgtgtgtatg gtaagagcac caaaaatcta ctctcttagc
12430      12440      12450      12460      12470      12480

aaatttctag tatacaatac aatgatattc gcaagaggca ggcaaatgcc taggcagaca
12490      12500      12510      12520      12530      12540

ggggcagggtc cctgggtgaaa cccacacctc aagccaaaga cagtttaaaag cctgaaagcc
12550      12560      12570      12580      12590      12600

aagctacaag tcaaattccac agactggatt gagaacctct cttcccatct ggtgcacttt
12610      12620      12630      12640      12650      12660

cctctgaatg atccctaccc ttcacctatt ttatatatac ctaccttcc caaattgggt
12670      12680      12690      12700      12710      12720

ttttacactg ccatgcccac tgagttgtgc ctttgtttta accttttctg cgtactcaca
12730      12740      12750      12760      12770      12780

aaccaatcac cactccccca ttctgagccc ataaaagccc cggagccagc cacactgaag
12790      12800      12810      12820      12830      12840

gggagaccac ccaacttcat gtggtggatc acctcatgg cccctctttg ctgagagctg
12850      12860      12870      12880      12890      12900

ttctgtcact taataaagct cttctctgcc cttctcacc ttcaattgtc agtgtaacct
12910      12920      12930      12940      12950      12960

cacttctctt ggattcagga caagaacaca agaattgctg aacataggta tgagctgtaa
12970      12980      12990      13000      13010      13020

cacaggtggg ctggggcact cctgggccag tcatgggggt agctagtgtg caaatgggca
13030      13040      13050      13060      13070      13080

gagtacctcc tgcagcaggc ctggggctga gcaaggccca gtttggggca tcgccagcca
13090      13100      13110      13120      13130      13140

gacacttcca gctggcaaaa gtgacagaaa aatcctgtat cagcaacact attaaactata
13150      13160      13170      13180      13190      13200

gtccacatgc tgcatatcag acctctagat ttattcatct gacaaaactg caactttgtt
13210      13220      13230      13240      13250      13260

tcctttttat ctacatctcc tgatctcccc ctacctctgg taaccacctt tctactctgt
13270      13280      13290      13300      13310      13320

ttatgtgtat tcagcatttt ttaagatgct gcatataaga aagatgtgga atacctttttt
13330      13340      13350      13360      13370      13380

tttctgtgtg cttatttcac tttagcgtaat gtcctctagg atcatcatgt tgcaaatagc
13390      13400      13410      13420      13430      13440

aggatctcct tcctttctaa ggctgaataa tattcgtgtg tgtgtgtgtg tgtgtgtgtg
13450      13460      13470      13480      13490      13500

tgtgtgtgtg tgtgtgtg
13510

```

Figure 10 (suite)

INTERNATIONAL SEARCH REPORT

International Application No

PCT/FR 97/02226

A. CLASSIFICATION OF SUBJECT MATTER

IPC 6	C12N15/53	C12N15/86	C12N15/11	C12N9/02	C12N5/10
	C12N1/21	C12Q1/26	C12Q1/68	C07K16/40	A61K38/44
	A61K39/395	A61K48/00	G01N33/50	G01N33/573	

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 C12N C12Q C07K A61K G01N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	MCCOMBIE R. ET AL.: "Localization of human flavin-containing monooxygenase genes FM02 and FM05 to chromosome 1q" GENOMICS, vol. 34, no. 3, 15 June 1996, pages 426-429, XP002039325 see the whole document, part. p.426, last paragraph ---	1,2,4,8
X	YUEH M. ET AL.: "AC U59453" EMBL DATABASE, 25 June 1996, HEIDELBERG, XP002039326 see the whole document --- -/--	1,4, 8-10, 12-14, 16,17

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search

31 March 1998

Date of mailing of the international search report

16/04/1998

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Kania, T

INTERNATIONAL SEARCH REPORT

International Application No

PCT/FR 97/02226

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	EP 0 712 932 A (SUMITOMO CHEMICAL CO) 22 May 1996 see the whole document ---	1,4,5, 8-18,20, 21,23, 24, 26-28, 60-63
X	PHILLIPS I. ET AL.: "The molecular biology of the flavin-containing monooxygenases of man" CHEMICO-BIOLOGICAL INTERACTIONS, vol. 96, no. 1, 28 April 1995, pages 17-32, XP002039327	1,2,4, 8-18,20, 21,23, 24,26
A	see the whole document, part. p. 21/22 *	60-63
X	---	
X	LARSEN-SU S. AND WILLIAMS D.: "Dietary indole-3-carbonil inhibits FMO activity and the expression of flavin-containing monooxygenase form 1 in rat liver and intestine" DRUG METABOLISM AND DISPOSITION, vol. 24, no. 9, September 1996, pages 927-931, XP002039328 see the whole document ---	47,49, 51,54,56
A	---	
A	SUNDEN S. ET AL.: "Fine mapping of the autosomal dominant juvenile open angle glaucoma (GLC1A) region and evaluation of candidate genes" GENOME RESEARCH, vol. 6, no. 9, September 1996, pages 862-869, XP002039329 cited in the application see the whole document ---	1-63
A	---	
A	LAWTON M. ET AL.: "A nomenclature for the mammalian flavin-containing monooxygenase gene family based on amino acid sequence identities" ARCHIVES OF BIOCHEMISTRY AND BIOPHYSICS, vol. 308, no. 1, January 1994, pages 254-257, XP002039330 cited in the application see the whole document ---	1-63
A	---	
A	GASSER R.: "The flavin-containing monooxygenase system" EXPERIMENTAL AND TOXICOLOGIC PATHOLOGY, vol. 48, no. 5, September 1996, pages 467-470, XP002039331 see the whole document ---	1-63
	-/--	

INTERNATIONAL SEARCH REPORT

International Application No

PCT/FR 97/02226

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	DOLPHIN C. ET AL.: "AC Y09267" EMBL DATABASE, 6 February 1997, HEIDELBERG, XP002060752 see the whole document -----	1,2,4, 8-10,16, 17

INTERNATIONAL SEARCH REPORT

information on patent family members

International Application No

PCT/FR 97/02226

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 0712932 A	22-05-96	JP 8140674 A	04-06-96
		CA 2163136 A	19-05-96

RAPPORT DE RECHERCHE INTERNATIONALE

Dema. Internationale No

PCT/FR 97/02226

A. CLASSEMENT DE L'OBJET DE LA DEMANDE

CIB 6	C12N15/53	C12N15/86	C12N15/11	C12N9/02	C12N5/10
	C12N1/21	C12Q1/26	C12Q1/68	C07K16/40	A61K38/44
	A61K39/395	A61K48/00	G01N33/50	G01N33/573	

Selon la classification internationale des brevets (CIB) ou à la fois selon la classification nationale et la CIB

B. DOMAINES SUR LESQUELS LA RECHERCHE A PORTE

Documentation minimale consultée (système de classification suivi des symboles de classement)

CIB 6 C12N C12Q C07K A61K G01N

Documentation consultée autre que la documentation minimale dans la mesure où ces documents relèvent des domaines sur lesquels a porté la recherche

Base de données électronique consultée au cours de la recherche internationale (nom de la base de données, et si cela est réalisable, termes de recherche utilisés)

C. DOCUMENTS CONSIDERES COMME PERTINENTS

Catégorie °	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées
X	MCCOMBIE R. ET AL.: "Localization of human flavin-containing monooxygenase genes FM02 and FM05 to chromosome 1q" GENOMICS, vol. 34, no. 3, 15 juin 1996, pages 426-429, XP002039325 * le document en entier, part. p.426, dernier par. *	1,2,4,8
X	YUEH M. ET AL.: "AC U59453" EMBL DATABASE, 25 juin 1996, HEIDELBERG, XP002039326 voir le document en entier --- -/-	1,4, 8-10, 12-14, 16,17



Voir la suite du cadre C pour la fin de la liste des documents



Les documents de familles de brevets sont indiqués en annexe

° Catégories spéciales de documents cités:

"A" document définissant l'état général de la technique, non considéré comme particulièrement pertinent

"E" document antérieur, mais publié à la date de dépôt international ou après cette date

"L" document pouvant jeter un doute sur une revendication de priorité ou cité pour déterminer la date de publication d'une autre citation ou pour une raison spéciale (telle qu'indiquée)

"O" document se référant à une divulgation orale, à un usage, à une exposition ou tous autres moyens

"P" document publié avant la date de dépôt international, mais postérieurement à la date de priorité revendiquée

"T" document ultérieur publié après la date de dépôt international ou la date de priorité et n'appartenant pas à l'état de la technique pertinent, mais cité pour comprendre le principe ou la théorie constituant la base de l'invention

"X" document particulièrement pertinent; l'invention revendiquée ne peut être considérée comme nouvelle ou comme impliquant une activité inventive par rapport au document considéré isolément

"Y" document particulièrement pertinent; l'invention revendiquée ne peut être considérée comme impliquant une activité inventive lorsque le document est associé à un ou plusieurs autres documents de même nature, cette combinaison étant évidente pour une personne du métier

"&" document qui fait partie de la même famille de brevets

Date à laquelle la recherche internationale a été effectivement achevée

31 mars 1998

Date d'expédition du présent rapport de recherche internationale

16/04/1998

Nom et adresse postale de l'administration chargée de la recherche internationale

Office Européen des Brevets, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Fonctionnaire autorisé

Kania, T

RAPPORT DE RECHERCHE INTERNATIONALE

Demande Internationale No
PCT/FR 97/02226

C.(suite) DOCUMENTS CONSIDERES COMME PERTINENTS

Catégorie	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées
X	EP 0 712 932 A (SUMITOMO CHEMICAL CO) 22 mai 1996 voir le document en entier ---	1,4,5, 8-18,20, 21,23, 24, 26-28, 60-63
X	PHILLIPS I. ET AL.: "The molecular biology of the flavin-containing monooxygenases of man" CHEMICO-BIOLOGICAL INTERACTIONS, vol. 96, no. 1, 28 avril 1995, pages 17-32, XP002039327	1,2,4, 8-18,20, 21,23, 24,26
A	* le document en entier, part. p. 21/22 *	60-63
X	---	
X	LARSEN-SU S. AND WILLIAMS D.: "Dietary indole-3-carbonil inhibits FMO activity and the expression of flavin-containing monooxygenase form 1 in rat liver and intestine" DRUG METABOLISM AND DISPOSITION, vol. 24, no. 9, septembre 1996, pages 927-931, XP002039328 voir le document en entier ---	47,49, 51,54,56
A	SUNDEN S. ET AL.: "Fine mapping of the autosomal dominant juvenile open angle glaucoma (GLC1A) region and evaluation of candidate genes" GENOME RESEARCH, vol. 6, no. 9, septembre 1996, pages 862-869, XP002039329 cité dans la demande voir le document en entier ---	1-63
A	LAWTON M. ET AL.: "A nomenclature for the mammalian flavin-containing monooxygenase gene family based on amino acid sequence identities" ARCHIVES OF BIOCHEMISTRY AND BIOPHYSICS, vol. 308, no. 1, janvier 1994, pages 254-257, XP002039330 cité dans la demande voir le document en entier ---	1-63
A	GASSER R.: "The flavin-containing monooxygenase system" EXPERIMENTAL AND TOXICOLOGIC PATHOLOGY, vol. 48, no. 5, septembre 1996, pages 467-470, XP002039331 voir le document en entier ---	1-63
	-/--	

RAPPORT DE RECHERCHE INTERNATIONALE

Demande Internationale No

PCT/FR 97/02226

C.(suite) DOCUMENTS CONSIDERES COMME PERTINENTS

Catégorie	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées
P,X	DOLPHIN C. ET AL.: "AC Y09267" EMBL DATABASE, 6 février 1997, HEIDELBERG, XP002060752 voir le document en entier -----	1,2,4, 8-10,16, 17

RAPPORT DE RECHERCHE INTERNATIONALE

Renseignements relatifs aux membres de familles de brevets

Demande Internationale No

PCT/FR 97/02226

Document brevet cité au rapport de recherche	Date de publication	Membre(s) de la famille de brevet(s)	Date de publication
EP 0712932 A	22-05-96	JP 8140674 A CA 2163136 A	04-06-96 19-05-96